

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 13:55:45 ; Search time 47.4522 Seconds
(without alignments)
1774.398 Million cell updates/sec

Title: US-09-864-675-4
Perfect score: 1574
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....KCPVGYTGDRCCQFAMVNFS 298

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		8								
Result		Query								
No.	Score	Match	Length	DB	ID	Description				

1	1574	100.0	298	5	AAU11636	Aau11636 Human Neu				
2	1547	98.3	754	2	AAW27536	Aaw27536 Rat cereb				
3	1505	95.6	330	5	AAU11635	Aau11635 Human Neu				
4	1505	95.6	422	5	ABB07894	Abb07894 Human neu				
5	1505	95.6	426	5	ABB07893	Abb07893 Human neu				
6	1478	93.9	330	2	AAW27537	Aaw27537 Rat cereb				
7	1470	93.4	860	2	AAW63700	Aaw63700 Receptor				
8	776	49.3	469	6	ABG71639	Abg71639 Human sec				
9	776	49.3	647	2	AAW48383	Aaw48383 Homo sapi				

10	776	49.3	647	6	ABG71644	Abg71644	Human thi
11	770	48.9	469	2	AAW48382	Aaw48382	Homo sapi
12	736	46.8	407	2	AAW48381	Aaw48381	Homo sapi
13	736	46.8	407	6	ABG71638	Abg71638	Human mem
14	716	45.5	181	2	AAW48380	Aaw48380	Mus muscu
15	716	45.5	181	6	ABG71637	Abg71637	Murine se
16	716	45.5	605	6	ABG71636	Abg71636	Murine me
17	711	45.2	605	2	AAW48379	Aaw48379	Mus muscu
18	545	34.6	422	4	AAG67901	Aag67901	Human neu
19	545	34.6	422	4	AAG67939	Aag67939	Human neu
20	544	34.6	422	2	AAR67258	Aar67258	Human gli
21	544	34.6	422	2	AAR96081	Aar96081	Glial gro
22	544	34.6	422	2	AAW09372	Aaw09372	Human GGF
23	544	34.6	422	2	AAW09371	Aaw09371	Human neu
24	544	34.6	422	2	AAR86628	Aar86628	Mature hG
25	544	34.6	422	2	AAR87466	Aar87466	Glial gro
26	543	34.5	422	2	AAR55654	Aar55654	GGF-II en
27	543	34.5	422	2	AAR46923	Aar46923	GGF-II en
28	542	34.4	418	5	ABJ00011	Abj00011	Human neu
29	542	34.4	418	5	ABJ00049	Abj00049	Human neu
30	529	33.6	422	2	AAR87467	Aar87467	Glial gro
31	506	32.1	782	4	AAB67751	Aab67751	Amino aci
32	504	32.0	139	2	AAW48388	Aaw48388	Undefined
33	504	32.0	139	6	ABG71645	Abg71645	Don-1 ass
34	495	31.4	855	4	AAB67757	Aab67757	Amino aci
35	476	30.2	342	4	AAB67754	Aab67754	Amino aci
36	473	30.1	182	2	AAW27538	Aaw27538	Human cer
37	473	30.1	323	4	AAB67753	Aab67753	Amino aci
38	471	29.9	317	4	AAB67752	Aab67752	Amino aci
39	375	23.8	204	4	AAG67902	Aag67902	Human neu
40	375	23.8	204	4	AAG67940	Aag67940	Human neu
41	375	23.8	204	5	ABJ00012	Abj00012	Human neu
42	375	23.8	204	5	ABJ00050	Abj00050	Human neu
43	362.5	23.0	257	2	AAR28538	Aar28538	GGF2BPP3.
44	362.5	23.0	257	2	AAR55690	Aar55690	GGF2BPP3.
45	362.5	23.0	257	2	AAR46897	Aar46897	GGF2BPP3.

ALIGNMENTS

RESULT 1

AAU11636

ID AAU11636 standard; protein; 298 AA.

XX

AC AAU11636;

XX

DT 12-MAR-2002 (first entry)

XX

DE Human Neuregulin-2beta, NRG-2beta.

XX

KW Human; neuregulin-2; NRG-2alpha; NRG-2beta; mitogenesis; cell survival;

KW cell growth; cell differentiation; erbB receptor; cardiomyopathy;

KW ischaemic damage; cardiac trauma; heart failure; atherosclerosis;

KW vascular lesion; vascular hypertension;

KW degenerative congenital vascular disease; myasthenia gravis;

KW neurodegenerative disorder; peripheral neuropathy;

KW sensory nerve fiber neuropathy; motor fiber neuropathy;
 KW sensory nerve fiber neuropathy; multiple sclerosis;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; nerve injury;
 KW Alzheimer's disease; Parkinson's disease; cerebellar ataxia;
 KW spinal cord injury; tumour; neurofibromatosis; transgenic animal.
 XX
 OS Homo sapiens.
 XX
 PN WO200189568-A1.
 XX
 PD 29-NOV-2001.
 XX
 PF 23-MAY-2001; 2001WO-US016896.
 XX
 PR 23-MAY-2000; 2000US-0206495P.
 XX
 PA (CENE-) CENES PHARM INC.
 XX
 PI Marchionni MA;
 XX
 DR WPI; 2002-097612/13.
 DR N-PSDB; AAS18020.
 XX
 PT Neuregulin-2 polypeptide and polynucleotide useful for treating multiple
 PT sclerosis, spinal muscular atrophy, nerve injury, Alzheimer's disease, by
 PT increasing mitogenesis, survival, growth or differentiation of a cell.
 XX
 PS Claim 53; Fig 9; 79pp; English.
 XX
 CC The invention relates to a substantially pure neuregulin (NRG)-2
 CC polypeptide comprising or consisting of a sequence for human NRG-2alpha
 CC or NRG-2beta (clone 2b7) and the polynucleotides encoding the. Also
 CC included are a vector expressing the protein, a host cell comprising the
 CC vector, a transgenic non-human animal transformed with the vector or
 CC having a knockout mutation in one or both NRG-2 alleles and an anti-NRG-2
 CC antibody. Analysis of mutations in NRG-2 in an individual is useful for
 CC diagnosing an increased likelihood of developing a NRG-2-related disease
 CC or condition in a test subject. NRG-2 is useful for increasing the
 CC mitogenesis, survival, growth or differentiation of a cell (e.g. a
 CC neuronal cell), where the cell expresses an erbB receptor. NRG-2 is
 CC useful for treating diseases and disorders such as cardiomyopathy
 CC (preferably degenerative congenital disease), ischaemic damage, cardiac
 CC trauma or heart failure or which has a condition affecting smooth muscle
 CC which include atherosclerosis, vascular lesion, vascular hypertension,
 CC and degenerative congenital vascular disease, myasthenia gravis, a
 CC neurodegenerative disorder, peripheral neuropathy, a sensory nerve fiber
 CC neuropathy, a motor fiber and a sensory nerve fiber neuropathy, multiple
 CC sclerosis, amyotrophic lateral sclerosis, spinal muscular atrophy, nerve
 CC injury, Alzheimer's disease, Parkinson's disease, cerebellar ataxia, and
 CC spinal cord injury. The antibody is useful for treatment of a tumour
 CC comprising inhibiting proliferation of a tumour cell preferably a glial
 CC tumour cell, for treating of neurofibromatosis by inhibiting glial cell
 CC mitogenesis. The present sequence represents NRG-2beta
 XX
 SQ Sequence 298 AA;

Query Match

100.0%; Score 1574; DB 5; Length 298;

Best Local Similarity 100.0%; Pred. No. 2.7e-105;

Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120

Qy    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCCQFAMVNFS 298
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCCQFAMVNFS 298
```

RESULT 2

AAW27536

ID AAW27536 standard; protein; 754 AA.

XX

AC AAW27536;

XX

DT 18-DEC-1997 (first entry)

XX

DE Rat cerebellum derived growth factor 1.

XX

KW Rat; cerebellum derived growth factor; CDGF1; screening; binding;

KW modulation; erbB type receptor; identification; indication; risk;

KW proliferation; differentiation; induction; neuron; hyperplasia;

KW stem cell culture; intracerebral graft; alleviation; repair;

KW behavioural defect; nervous system; central; peripheral; nerve;

KW prosthesis; damage; entubulation; cell survival; treatment; injury;

KW trauma; ischaemia; ischemia; stroke; infection; disorder; inflammation;

KW neurodegeneration; disease; Parkinson's; Huntingdon's;

KW amyotrophic lateral sclerosis; sensory; retina;

KW spinocerebellar degeneration; multiple sclerosis; neoplasia;

KW amalignant glioma; medulloblastoma; neuroectodermal tumour.

XX

OS Rattus rattus.

XX

FH Key Location/Qualifiers

FT Peptide 1. .23

FT /label= sig_peptide

FT Peptide 24. .754

FT /label= mat_peptide

FT Region 55

FT /note= "potential N-glycosylation site"

FT Domain 158. .228

FT /label= immunoglobulin_like_domain

FT Region 186
 FT /note= "potential N-glycosylation site"
 FT Domain 252. .297
 FT /label= epidermal_growth_factor_like_domain
 FT Region 253
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 254
 FT /note= "potential N-glycosylation site"
 FT Region 261
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 267
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 278
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 280
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 289
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 296
 FT /note= "potential N-glycosylation site"
 FT Cleavage-site 314. .315
 FT /label= potential_proteolytic_site
 FT Domain 316. .338
 FT /label= putative_transmembrane_domain
 XX
 PN WO9709425-A1.
 XX
 PD 13-MAR-1997.
 XX
 PF 09-SEP-1996; 96WO-US014484.
 XX
 PR 08-SEP-1995; 95US-00525864.
 XX
 PA (HARD) HARVARD COLLEGE.
 PA (STRD) UNIV LELAND S STANFORD.
 XX
 PI Chang H;
 XX
 DR WPI; 1997-192900/17.
 DR N-PSDB; AAT87922.
 XX
 PT Rat and human cerebellum-derived growth factors - used in the treatment
 PT of neuronal injury and proliferative disorders.
 XX
 PS Claim 1; Page 63-66; 94pp; English.
 XX
 CC The present sequence is rat cerebellum derived growth factor 1 (CDGF1),
 CC which can be used to screen for modulators of CDGF binding to erbB type
 CC receptors. Identification of a modification or mutation in a CDGF gene,
 CC or aberrant expression of a CDGF gene or levels of soluble CDGF may be
 CC used to indicate the risk of unwanted cell proliferation or

CC differentiation. CDGF may be used to induce neuronal differentiation in
CC stem cell culture, and maintain the integrity of a terminally
CC differentiated neuronal cell culture, e.g. useful for intracerebral
CC grafting to alleviate behavioural defects. CDGF may also be used in nerve
CC protheses to repair central and peripheral nerve damage, especially where
CC a crushed or severed axon is entubulated by a prosthetic. CDGF may also
CC be used to enhance neuronal cell survival in the central or peripheral
CC nervous system, to treat neurological conditions associated with nervous
CC system injury, e.g. traumatic, chemical or vasal injury and deficits such
CC as ischaemia resulting from stroke, infectious/inflammatory and tumour
CC induced injury, chronic neurodegenerative disease including Parkinson's
CC and Huntingdon's, amyotrophic lateral sclerosis, spinocerebellar
CC degeneration, chronic immunological disease of the nervous system
CC including multiple sclerosis, disorders of the sensory neurons and
CC degenerative diseases of the retina. CDGF may also be used to treat
CC neoplastic or hyperplastic transformations, particularly of the central
CC nervous system, e.g. amalignant gliomas, medulloblastomas and
CC neuroectodermal tumours

XX

SQ Sequence 754 AA;

Query Match 98.3%; Score 1547; DB 2; Length 754;
Best Local Similarity 97.7%; Pred. No. 6.9e-103;
Matches 291; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
          |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |:| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db    121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
          |||
Db    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLSCKCPVGYTGDRQQFAMVNFS 298
          |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLSCKCPVGYTGDRQQFAMVNFS 298
```

RESULT 3

AAU11635

ID AAU11635 standard; protein; 330 AA.

XX

AC AAU11635;

XX

DT 12-MAR-2002 (first entry)

XX

DE Human Neuregulin-2alpha, NRG-2alpha.

XX

KW Human; neuregulin-2; NRG-2alpha; NRG-2beta; mitogenesis; cell survival;

KW cell growth; cell differentiation; erbB receptor; cardiomyopathy;
 KW ischaemic damage; cardiac trauma; heart failure; atherosclerosis;
 KW vascular lesion; vascular hypertension;
 KW degenerative congenital vascular disease; myasthenia gravis;
 KW neurodegenerative disorder; peripheral neuropathy;
 KW sensory nerve fiber neuropathy; motor fiber neuropathy;
 KW sensory nerve fiber neuropathy; multiple sclerosis;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; nerve injury;
 KW Alzheimer's disease; Parkinson's disease; cerebellar ataxia;
 KW spinal cord injury; tumour; neurofibromatosis; transgenic animal.
 XX
 OS Homo sapiens.
 XX
 PN WO200189568-A1.
 XX
 PD 29-NOV-2001.
 XX
 PF 23-MAY-2001; 2001WO-US016896.
 XX
 PR 23-MAY-2000; 2000US-0206495P.
 XX
 PA (CENE-) CENES PHARM INC.
 XX
 PI Marchionni MA;
 XX
 DR WPI; 2002-097612/13.
 DR N-PSDB; AAS18019.
 XX
 PT Neuregulin-2 polypeptide and polynucleotide useful for treating multiple
 PT sclerosis, spinal muscular atrophy, nerve injury, Alzheimer's disease, by
 PT increasing mitogenesis, survival, growth or differentiation of a cell.
 XX
 PS Claim 53; Fig 7; 79pp; English.
 XX
 CC The invention relates to a substantially pure neuregulin (NRG)-2
 CC polypeptide comprising or consisting of a sequence for human NRG-2alpha
 CC or NRG-2beta (clone 2b7) and the polynucleotides encoding the. Also
 CC included are a vector expressing the protein, a host cell comprising the
 CC vector, a transgenic non-human animal transformed with the vector or
 CC having a knockout mutation in one or both NRG-2 alleles and an anti-NRG-2
 CC antibody. Analysis of mutations in NRG-2 in an individual is useful for
 CC diagnosing an increased likelihood of developing a NRG-2-related disease
 CC or condition in a test subject. NRG-2 is useful for increasing the
 CC mitogenesis, survival, growth or differentiation of a cell (e.g. a
 CC neuronal cell), where the cell expresses an erbB receptor. NRG-2 is
 CC useful for treating diseases and disorders such as cardiomyopathy
 CC (preferably degenerative congenital disease), ischaemic damage, cardiac
 CC trauma or heart failure or which has a condition affecting smooth muscle
 CC which include atherosclerosis, vascular lesion, vascular hypertension,
 CC and degenerative congenital vascular disease, myasthenia gravis, a
 CC neurodegenerative disorder, peripheral neuropathy, a sensory nerve fiber
 CC neuropathy, a motor fiber and a sensory nerve fiber neuropathy, multiple
 CC sclerosis, amyotrophic lateral sclerosis, spinal muscular atrophy, nerve
 CC injury, Alzheimer's disease, Parkinson's disease, cerebellar ataxia, and
 CC spinal cord injury. The antibody is useful for treatment of a tumour
 CC comprising inhibiting proliferation of a tumour cell preferably a glial
 CC tumour cell, for treating of neurofibromatosis by inhibiting glial cell

CC mitogenesis. The present sequence represents NRG-2alpha
XX
SQ Sequence 330 AA;

Query Match 95.6%; Score 1505; DB 5; Length 330;
Best Local Similarity 98.6%; Pred. No. 2.8e-100;
Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120
          |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120

Qy    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |||
Db    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
          |||
Db    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRG 289
          |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289
```

RESULT 4

ABB07894

ID ABB07894 standard; protein; 422 AA.

XX

AC ABB07894;

XX

DT 03-JUL-2002 (first entry)

XX

DE Human neuregulin 2 isoform 6.

XX

KW Human; MUC1; mucin; glycoprotein; cytostatic; cancer; tumour; ECD;
KW extracellular domain; neuregulin 2; isoform.

XX

OS Homo sapiens.

XX

PN WO200222685-A2.

XX

PD 21-MAR-2002.

XX

PF 11-SEP-2001; 2001WO-US028548.

XX

PR 11-SEP-2000; 2000US-0231841P.

XX

PA (KUFE/) KUFE D W.

PA (OHNO/) OHNO T.

XX

PI Kufe DW, Ohno T;

XX

DR WPI; 2002-339864/37.

XX

PT Use of a mucin glycoprotein (MUC1) extracellular domain antagonist for
PT manufacturing a medicant that inhibits the proliferation of MUC-1
PT expressing cancer cells and that can treat cancers and reduce tumor
PT growth.

XX

PS Claim 6; Page 56-58; 74pp; English.

XX

CC The invention relates to the use of a MUC1 (mucin glycoprotein)
CC extracellular domain (ECD) antagonist for the manufacture of a medicant
CC to inhibit the proliferation of MUC-1 expressing cancer cells. MUC1 ECD
CC antagonists (optionally combined with a pharmaceutical carrier) can be
CC administered to inhibit proliferation of MUC1-expressing cancer cells,
CC useful to treat cancers e.g. skin cancer, prostate cancer and leukemia,
CC especially in humans. The method may also be combined with administration
CC of a chemotherapeutic agent (e.g. an alkylating agent, topoisomerase etc)
CC or radiation to treat cancer, especially to reduce tumour growth. The
CC polypeptides are also useful in screening to identify MUC1 ECD
CC antagonists. The present sequence represents a human neuregulin 2 isoform
CC 6, a fragment of which can bind to MUC1/ECD

XX

SQ Sequence 422 AA;

Query Match 95.6%; Score 1505; DB 5; Length 422;

Best Local Similarity 98.6%; Pred. No. 3.7e-100;

Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      93 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
          |||
Db     153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212

Qy     121 PLDTNGKNLKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |||
Db     213 PLDTNGKNLKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272

Qy     181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
          |||
Db     273 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 332

Qy     241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRC 289
          |||
Db     333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 381
```

RESULT 5

ABB07893

ID ABB07893 standard; protein; 426 AA.

XX

AC ABB07893;

XX

DT 03-JUL-2002 (first entry)

XX

DE Human neuregulin 2 isoform 5.
 XX
 KW Human; MUC1; mucin; glycoprotein; cytostatic; cancer; tumour; ECD;
 KW extracellular domain; neuregulin 2; isoform.
 XX
 OS Homo sapiens.
 XX
 PN WO200222685-A2.
 XX
 PD 21-MAR-2002.
 XX
 PF 11-SEP-2001; 2001WO-US028548.
 XX
 PR 11-SEP-2000; 2000US-0231841P.
 XX
 PA (KUFE/) KUFE D W.
 PA (OHNO/) OHNO T.
 XX
 PI Kufe DW, Ohno T;
 XX
 DR WPI; 2002-339864/37.
 XX
 PT Use of a mucin glycoprotein (MUC1) extracellular domain antagonist for
 PT manufacturing a medicant that inhibits the proliferation of MUC-1
 PT expressing cancer cells and that can treat cancers and reduce tumor
 PT growth.
 XX
 PS Claim 6; Page 53-55; 74pp; English.
 XX
 CC The invention relates to the use of a MUC1 (mucin glycoprotein)
 CC extracellular domain (ECD) antagonist for the manufacture of a medicant
 CC to inhibit the proliferation of MUC-1 expressing cancer cells. MUC1 ECD
 CC antagonists (optionally combined with a pharmaceutical carrier) can be
 CC administered to inhibit proliferation of MUC1-expressing cancer cells,
 CC useful to treat cancers e.g. skin cancer, prostate cancer and leukemia,
 CC especially in humans. The method may also be combined with administration
 CC of a chemotherapeutic agent (e.g. an alkylating agent, topoisomerase etc)
 CC or radiation to treat cancer, especially to reduce tumour growth. The
 CC polypeptides are also useful in screening to identify MUC1 ECD
 CC antagonists. The present sequence represents a human neuregulin 2 isoform
 CC 5, a fragment of which can bind to MUC1/ECD
 XX
 SQ Sequence 426 AA;

Query Match 95.6%; Score 1505; DB 5; Length 426;
 Best Local Similarity 98.6%; Pred. No. 3.7e-100;
 Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
 |||
 Db 93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152
 Qy 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
 |||
 Db 153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212
 Qy 121 PLDTNGKNLKKKEVGKILCTDCATRPKLKMKKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

```

                |||
Db      213 PLDTNGKNLKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272
                |||
Qy      181 DGKELNRSRDIRIKYGNRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVS 240
                |||
Db      273 DGKELNRSRDIRIKYGNRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVS 332
                |||
Qy      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRG 289
                |||
Db      333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 381

```

RESULT 6

AAW27537

ID AAW27537 standard; protein; 330 AA.

XX

AC AAW27537;

XX

DT 18-DEC-1997 (first entry)

XX

DE Rat cerebellum derived growth factor 2.

XX

KW Rat; cerebellum derived growth factor; CDGF2; screening; binding;
KW modulation; erbB type receptor; identification; indication; risk;
KW proliferation; differentiation; induction; neuron; hyperplasia;
KW stem cell culture; intracerebral graft; alleviation; repair;
KW behavioural defect; nervous system; central; peripheral; nerve;
KW prosthesis; damage; entubulation; cell survival; treatment; injury;
KW trauma; ischaemia; ischemia; stroke; infection; disorder; inflammation;
KW neurodegeneration; disease; Parkinson's; Huntington's;
KW amyotrophic lateral sclerosis; sensory; retina;
KW spinocerebellar degeneration; multiple sclerosis; neoplasia;
KW amalignant glioma; medulloblastoma; neuroectodermal tumour.

XX

OS Rattus rattus.

XX

FH	Key	Location/Qualifiers
FT	Peptide	1. .23
FT		/label= sig_peptide
FT	Peptide	24. .330
FT		/label= mat_peptide
FT	Region	55
FT		/note= "potential N-glycosylation site"
FT	Domain	158. .228
FT		/label= immunoglobulin_like_domain
FT	Region	186
FT		/note= "potential N-glycosylation site"
FT	Domain	252. .297
FT		/label= epidermal_growth_factor_like_domain
FT	Region	253
FT		/note= "characteristic cysteine of epidermal growth factor like domain"
FT	Region	254
FT		/note= "potential N-glycosylation site"
FT	Region	261
FT		/note= "characteristic cysteine of epidermal growth factor like domain"

FT Region 267
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 278
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 280
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 FT Region 289
 FT /note= "characteristic cysteine of epidermal growth
 FT factor like domain"
 XX
 PN WO9709425-A1.
 XX
 PD 13-MAR-1997.
 XX
 PF 09-SEP-1996; 96WO-US014484.
 XX
 PR 08-SEP-1995; 95US-00525864.
 XX
 PA (HARD) HARVARD COLLEGE.
 PA (STRD) UNIV LELAND S STANFORD.
 XX
 PI Chang H;
 XX
 DR WPI; 1997-192900/17.
 DR N-PSDB; AAT87923.
 XX
 PT Rat and human cerebellum-derived growth factors - used in the treatment
 PT of neuronal injury and proliferative disorders.
 XX
 PS Claim 1; Page 70-71; 94pp; English.
 XX
 CC The present sequence is rat cerebellum derived growth factor 2 (CDGF2),
 CC which can be used to screen for modulators of CDGF binding to erbB type
 CC receptors. Identification of a modification or mutation in a CDGF gene,
 CC or aberrant expression of a CDGF gene or levels of soluble CDGF may be
 CC used to indicate the risk of unwanted cell proliferation or
 CC differentiation. CDGF may be used to induce neuronal differentiation in
 CC stem cell culture, and maintain the integrity of a terminally
 CC differentiated neuronal cell culture, e.g. useful for intracerebral
 CC grafting to alleviate behavioural defects. CDGF may also be used in nerve
 CC protheses to repair central and peripheral nerve damage, especially where
 CC a crushed or severed axon is entubulated by a prosthetic. CDGF may also
 CC be used to enhance neuronal cell survival in the central or peripheral
 CC nervous system, to treat neurological conditions associated with nervous
 CC system injury, e.g. traumatic, chemical or vasal injury and deficits such
 CC as ischaemia resulting from stroke, infectious/inflammatory and tumour
 CC induced injury, chronic neurodegenerative disease including Parkinson's
 CC and Huntingdon's, amyotrophic lateral sclerosis, spinocerebellar
 CC degeneration, chronic immunological disease of the nervous system
 CC including multiple sclerosis, disorders of the sensory neurons and
 CC degenerative diseases of the retina. CDGF may also be used to treat
 CC neoplastic or hyperplastic transformations, particularly of the central
 CC nervous system, e.g. amalignant gliomas, medulloblastomas and
 CC neuroectodermal tumours

XX

SQ Sequence 330 AA;

Query Match 93.9%; Score 1478; DB 2; Length 330;

Best Local Similarity 96.2%; Pred. No. 2.4e-98;

Matches 278; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120
        |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFE 120

Qy    121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
        |::|
Db    121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
        |||
Db    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRG 289
        |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289
```

RESULT 7

AAW63700

ID AAW63700 standard; protein; 860 AA.

XX

AC AAW63700;

XX

DT 29-SEP-1998 (first entry)

XX

DE Receptor type tyrosine kinase ErbB ligand.

XX

KW Receptor type tyrosine kinase ErbB; ligand; diagnostic agent;
KW nervous disease; cancer.

XX

OS Rattus sp.

XX

PN JP10179166-A.

XX

PD 07-JUL-1998.

XX

PF 25-DEC-1996; 96JP-00356998.

XX

PR 25-DEC-1996; 96JP-00356998.

XX

PA (HIGA/) HIGASHIYAMA S.

XX

DR WPI; 1998-430952/37.

DR N-PSDB; AAV43674.

XX

PT Gene coding the ligand of the tyrosine kinase ErbB receptor - useful for

XX
 FH Key Location/Qualifiers
 FT Misc-difference 14
 FT /note= "Encoded by AA"
 XX
 PN US2002127594-A1.
 XX
 PD 12-SEP-2002.
 XX
 PF 12-MAR-2002; 2002US-00096241.
 XX
 PR 22-JUN-2000; 2000US-00599789.
 XX
 PA (GEAR/) GEARING D P.
 PA (BUSF/) BUSFIELD S J.
 XX
 PI Gearing DP, Busfield SJ;
 XX
 DR WPI; 2003-039584/03.
 DR N-PSDB; ABS56036.
 XX
 PT Novel Don-1 polypeptide useful for stimulating proliferation of cells,
 PT for identifying proteins that interact with Don-1, and for regulating
 PT tumor formation and progression in brain.
 XX
 PS Claim 25; Fig 4; 66pp; English.
 XX
 CC The present invention relates to the isolation of a novel gene called Don
 CC -1, and alternate splice variants of Don-1, which are related to
 CC epidermal growth factors (EGF) such as neuregulins. Don-1 polypeptides
 CC are glycoprotein ligands. Both murine and human Don-1 sequences are
 CC cloned. The mouse Don-1 gene maps to chromosome 18. Don-1 polypeptides
 CC are useful for stimulating proliferation of a cell. Antibodies to Don-1
 CC polypeptides are useful for detecting Don-1 in a sample. The Don-1
 CC polypeptides are useful for treating and diagnosing cell proliferative
 CC disorders and play a role in the proliferation of carcinomas e.g.
 CC adenocarcinoma, myeloma, in cell differentiation, proliferation and
 CC survival. The polypeptides are also useful for inhibiting proliferation
 CC of adenocarcinoma cells, for stimulating the proliferation of cells such
 CC as epithelial cells to promote wound healing, for identifying proteins
 CC that interact with Don-1, and for regulating tumour formation and
 CC progression in the brain. The polynucleotide sequences encoding Don-1 may
 CC be used in gene therapy. The present sequence represents human second
 CC splice variant of Don-1
 XX
 SQ Sequence 469 AA;

Query Match 49.3%; Score 776; DB 6; Length 469;
 Best Local Similarity 97.3%; Pred. No. 1.2e-47;
 Matches 144; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
 |
 Db 31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90
 |
 Qy 202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
 |

Db 91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy 262 VNGGVCYYIEGINQLSCKCPVGYTGDRC 289
 ||||| : ||

Db 151 VNGGVCYYIEGINQLSCKCPNGFFGQRC 178

RESULT 9

AAW48383

ID AAW48383 standard; protein; 647 AA.

XX

AC AAW48383;

XX

DT 17-AUG-1998 (first entry)

XX

DE Homo sapiens don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;
 KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;
 KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;
 KW wound healing; transmembrane.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Domain	54. .108
----	--------	----------

FT		/note= "Ig domain"
----	--	--------------------

FT	Domain	142. .178
----	--------	-----------

FT		/note= "EGF domain"
----	--	---------------------

FT	Domain	203. .225
----	--------	-----------

FT		/note= "transmembrane domain"
----	--	-------------------------------

FT	Domain	226. .647
----	--------	-----------

FT		/note= "cytoplasmic domain"
----	--	-----------------------------

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.

XX

PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 1998-169084/15.

DR N-PSDB; AAV17816.

XX

PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas
 PT and adenocarcinoma(s), and for wound healing.

XX

PS Claim 25; Fig 7; 121pp; English.

XX

CC The sequence is that encoded by a human don-1 gene splice variant. Don-1
 CC polypeptides stimulate proliferation of epithelial cells and thus are

CC implicated in melanomas and adenocarcinomas in which epithelial cells
 CC proliferate out of control. Compounds that interfere with don-1 mediated
 CC cell proliferation can be used in the treatment of tumours such as
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial
 CC cell proliferation, e.g. for wound healing

XX

SQ Sequence 647 AA;

Query Match 49.3%; Score 776; DB 2; Length 647;

Best Local Similarity 97.3%; Pred. No. 1.8e-47;

Matches 144; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90
 Qy 202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 261
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 150
 Qy 262 VNGGVCYYIEGINQLSCKCPVGYTGDR 289
 |||||||||||||||||||||:|
 Db 151 VNGGVCYYIEGINQLSCKCPNGFFGQRC 178

RESULT 10

ABG71644

ID ABG71644 standard; protein; 647 AA.

XX

AC ABG71644;

XX

DT 14-JAN-2003 (first entry)

XX

DE Human third splice variant of Don-1.

XX

KW Human; Don-1; epidermal growth factor; EGF; neuregulin;

KW glycoprotein ligand; cell proliferation; cell proliferative disorder;

KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;

KW cell survival; epithelial cell; wound healing; tumour formation; brain;

KW vulnerary; cytostatic.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 14

FT /note= "Encoded by AA"

FT Misc-difference 310

FT /note= "Encoded by AGC"

XX

PN US2002127594-A1.

XX

PD 12-SEP-2002.

XX

PF 12-MAR-2002; 2002US-00096241.

XX

XX

PA

XX

XX

DR

XX

XX

XX

XX

Best Local Similarity 97.3%; Pred. No. 1.8e-47;

Db

Dh

10

AAW48382

XX

AC AAW48382;
 XX
 DT 17-AUG-1998 (first entry)
 XX
 DE Homo sapiens don-1 polypeptide.
 XX
 KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;
 KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;
 KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;
 KW wound healing; transmembrane.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 54. .108
 FT /note= "Ig domain"
 FT Domain 142. .178
 FT /note= "EGF domain"
 FT Domain 203. .225
 FT /note= "transmembrane domain"
 FT Domain 226. .469
 FT /note= "cytoplasmic domain"
 XX
 PN WO9807736-A1.
 XX
 PD 26-FEB-1998.
 XX
 PF 18-AUG-1997; 97WO-US014585.
 XX
 PR 19-AUG-1996; 96US-00699591.
 PR 19-NOV-1996; 96US-00753007.
 XX
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 XX
 PI Gearing DP, Busfield SJ;
 XX
 DR WPI; 1998-169084/15.
 DR N-PSDB; AAV17815.
 XX
 PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas
 PT and adenocarcinoma(s), and for wound healing.
 XX
 PS Claim 25; Fig 4; 121pp; English.
 XX
 CC The sequence is that encoded by a human don-1 gene splice variant. Don-1
 CC polypeptides stimulate proliferation of epithelial cells and thus are
 CC implicated in melanomas and adenocarcinomas in which epithelial cells
 CC proliferate out of control. Compounds that interfere with don-1 mediated
 CC cell proliferation can be used in the treatment of tumours such as
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial
 CC cell proliferation, e.g. for wound healing
 XX
 SQ Sequence 469 AA;

Query Match 48.9%; Score 770; DB 2; Length 469;

Best Local Similarity 96.6%; Pred. No. 3.4e-47;
Matches 143; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy      142 ATRPKLKKMKSQTGQVGEEKQLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEEKQLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPVGYTGDRG 289
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFAQRC 178
```

RESULT 12

AAW48381

ID AAW48381 standard; protein; 407 AA.

XX

AC AAW48381;

XX

DT 17-AUG-1998 (first entry)

XX

DE Homo sapiens don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;

KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;

KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;

KW wound healing; transmembrane.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain 16. .70

FT /note= "Ig domain"

FT Domain 104. .140

FT /note= "EGF domain"

FT Region 157. .164

FT /note= "juxtamembrane region"

FT Domain 173. .195

FT /note= "transmembrane domain"

FT Domain 196. .407

FT /note= "cytoplasmic domain"

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.

XX

PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX

PI Gearing DP, Busfield SJ;

XX
 DR WPI; 1998-169084/15.
 DR N-PSDB; AAV17814.
 XX
 PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas
 PT and adenocarcinoma(s), and for wound healing.
 XX
 PS Claim 25; Fig 3; 121pp; English.
 XX
 CC The sequence is that encoded by a human don-1 gene splice variant. Don-1
 CC polypeptides stimulate proliferation of epithelial cells and thus are
 CC implicated in melanomas and adenocarcinomas in which epithelial cells
 CC proliferate out of control. Compounds that interfere with don-1 mediated
 CC cell proliferation can be used in the treatment of tumours such as
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial
 CC cell proliferation, e.g. for wound healing
 XX
 SQ Sequence 407 AA;

Query Match 46.8%; Score 736; DB 2; Length 407;
 Best Local Similarity 97.1%; Pred. No. 8e-45;
 Matches 136; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 60
 Qy 210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120
 Qy 270 IEGINQLSCKCPVGYTGDR 289
 ||||||||| | : | ||
 Db 121 IEGINQLSCKCPNGFFGQRC 140

RESULT 13

ABG71638

ID ABG71638 standard; protein; 407 AA.

XX

AC ABG71638;

XX

DT 14-JAN-2003 (first entry)

XX

DE Human membrane-bound splice variant of Don-1.

XX

KW Human; Don-1; epidermal growth factor; EGF; neuregulin;

KW glycoprotein ligand; cell proliferation; cell proliferative disorder;

KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;

KW cell survival; epithelial cell; wound healing; tumour formation; brain;

KW vulnerary; cytostatic.

XX

OS Homo sapiens.

XX

PN US2002127594-A1.

XX
 PD 12-SEP-2002.
 XX
 PF 12-MAR-2002; 2002US-00096241.
 XX
 PR 22-JUN-2000; 2000US-00599789.
 XX
 PA (GEAR/) GEARING D P.
 PA (BUSF/) BUSFIELD S J.
 XX
 PI Gearing DP, Busfield SJ;
 XX
 DR WPI; 2003-039584/03.
 DR N-PSDB; ABS56035.
 XX
 PT Novel Don-1 polypeptide useful for stimulating proliferation of cells,
 PT for identifying proteins that interact with Don-1, and for regulating
 PT tumor formation and progression in brain.
 XX
 PS Claim 25; Fig 3; 66pp; English.
 XX
 CC The present invention relates to the isolation of a novel gene called Don
 CC -1, and alternate splice variants of Don-1, which are related to
 CC epidermal growth factors (EGF) such as neuregulins. Don-1 polypeptides
 CC are glycoprotein ligands. Both murine and human Don-1 sequences are
 CC cloned. The mouse Don-1 gene maps to chromosome 18. Don-1 polypeptides
 CC are useful for stimulating proliferation of a cell. Antibodies to Don-1
 CC polypeptides are useful for detecting Don-1 in a sample. The Don-1
 CC polypeptides are useful for treating and diagnosing cell proliferative
 CC disorders and play a role in the proliferation of carcinomas e.g.
 CC adenocarcinoma, myeloma, in cell differentiation, proliferation and
 CC survival. The polypeptides are also useful for inhibiting proliferation
 CC of adenocarcinoma cells, for stimulating the proliferation of cells such
 CC as epithelial cells to promote wound healing, for identifying proteins
 CC that interact with Don-1, and for regulating tumour formation and
 CC progression in the brain. The polynucleotide sequences encoding Don-1 may
 CC be used in gene therapy. The present sequence represents human membrane-
 CC bound splice variant of Don-1
 XX
 SQ Sequence 407 AA;

Query Match 46.8%; Score 736; DB 6; Length 407;
 Best Local Similarity 97.1%; Pred. No. 8e-45;
 Matches 136; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
 |
 Db 1 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 60

 Qy 210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
 |
 Db 61 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

 Qy 270 IEGINQLSCKCPVGYTGDR 289
 |
 Db 121 IEGINQLSCKCPNGFFGQRC 140

RESULT 14

AAW48380

ID AAW48380 standard; protein; 181 AA.

XX

AC AAW48380;

XX

DT 17-AUG-1998 (first entry)

XX

DE Mus musculus don-1 polypeptide.

XX

KW Murine; don-1 gene; melanoma; treatment; adenocarcinoma; epithelial cell;
 KW proliferation; stimulation; treatment; tumours; skin; oesophagus; lung;
 KW breast; liver; pancreas; colon; prostate; gastrointestinal tract; uterus;
 KW wound healing; transmembrane.

XX

OS Mus musculus.

XX

FH Key Location/Qualifiers

FT Domain 104. .140

FT /note= "EGF domain"

XX

PN WO9807736-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97WO-US014585.

XX

PR 19-AUG-1996; 96US-00699591.

PR 19-NOV-1996; 96US-00753007.

XX

PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 1998-169084/15.

DR N-PSDB; AAV17813.

XX

PT Mouse and human don-1 polypeptide(s) - useful for treatment of melanomas
 PT and adenocarcinoma(s), and for wound healing.

XX

PS Claim 25; Fig 2; 121pp; English.

XX

CC The sequence is that encoded by a murine don-1 gene splice variant. Don-1
 CC polypeptides stimulate proliferation of epithelial cells and thus are
 CC implicated in melanomas and adenocarcinomas in which epithelial cells
 CC proliferate out of control. Compounds that interfere with don-1 mediated
 CC cell proliferation can be used in the treatment of tumours such as
 CC melanomas and adenocarcinomas of the skin, oesophagus, lung, breast,
 CC liver, pancreas, gastrointestinal tract, colon, prostate or uterus.
 CC Alternatively, don-1 polypeptides can be used to stimulate epithelial
 CC cell proliferation, e.g. for wound healing

XX

SQ Sequence 181 AA;

Query Match 45.5%; Score 716; DB 2; Length 181;

Best Local Similarity 94.3%; Pred. No. 8.5e-44;

Matches 132; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```
Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db      1   MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRRLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPVGYTGDRC 289
          ||||| ||: ||
Db      121 IEGINQLSCKCPNGFFGQRC 140
```

RESULT 15

ABG71637

ID ABG71637 standard; protein; 181 AA.

XX

AC ABG71637;

XX

DT 14-JAN-2003 (first entry)

XX

DE Murine secreted splice variant of Don-1.

XX

KW Murine; Don-1; epidermal growth factor; EGF; neuregulin; mouse;
KW glycoprotein ligand; cell proliferation; cell proliferative disorder;
KW carcinoma; adenocarcinoma cell; myeloma; cell differentiation;
KW cell survival; epithelial cell; wound healing; tumour formation; brain;
KW vulnerary; cytostatic.

XX

OS Mus sp.

XX

PN US2002127594-A1.

XX

PD 12-SEP-2002.

XX

PF 12-MAR-2002; 2002US-00096241.

XX

PR 22-JUN-2000; 2000US-00599789.

XX

PA (GEAR/) GEARING D P.

PA (BUSF/) BUSFIELD S J.

XX

PI Gearing DP, Busfield SJ;

XX

DR WPI; 2003-039584/03.

DR N-PSDB; ABS56034.

XX

PT Novel Don-1 polypeptide useful for stimulating proliferation of cells,
PT for identifying proteins that interact with Don-1, and for regulating
PT tumor formation and progression in brain.

XX

PS Claim 25; Fig 2; 66pp; English.

XX

CC The present invention relates to the isolation of a novel gene called Don
CC -1, and alternate splice variants of Don-1, which are related to

CC epidermal growth factors (EGF) such as neuregulins. Don-1 polypeptides
 CC are glycoprotein ligands. Both murine and human Don-1 sequences are
 CC cloned. The mouse Don-1 gene maps to chromosome 18. Don-1 polypeptides
 CC are useful for stimulating proliferation of a cell. Antibodies to Don-1
 CC polypeptides are useful for detecting Don-1 in a sample. The Don-1
 CC polypeptides are useful for treating and diagnosing cell proliferative
 CC disorders and play a role in the proliferation of carcinomas e.g.
 CC adenocarcinoma, myeloma, in cell differentiation, proliferation and
 CC survival. The polypeptides are also useful for inhibiting proliferation
 CC of adenocarcinoma cells, for stimulating the proliferation of cells such
 CC as epithelial cells to promote wound healing, for identifying proteins
 CC that interact with Don-1, and for regulating tumour formation and
 CC progression in the brain. The polynucleotide sequences encoding Don-1 may
 CC be used in gene therapy. The present sequence represents murine secreted
 CC splice variant of Don-1
 XX

SQ Sequence 181 AA;

Query Match 45.5%; Score 716; DB 6; Length 181;
 Best Local Similarity 94.3%; Pred. No. 8.5e-44;
 Matches 132; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKNSRLQFNKV 209
 |||||:||||||||||||||||||||||||||||||||||||| |||||
 Db 1 MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKNSRLQFNKV 60
 Qy 210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
 :|||||||||||||||||||||:|||||||||||||||||||||
 Db 61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120
 Qy 270 IEGINQLSCKCPVGYTGDR 289
 ||||| ||| |: ||
 Db 121 IEGINQLSCKCPNGFFGQRC 140

Search completed: August 17, 2004, 14:10:49
 Job time : 48.4522 secs

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:09:05 ; Search time 14.7102 Seconds
 (without alignments)
 1045.842 Million cell updates/sec

Title: US-09-864-675-4
 Perfect score: 1574
 Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....KCPVGYTGDRCQQFAMVNFS 298

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued_Patents_AA:*
 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	1547	98.3	754	2	US-08-525-864A-2	Sequence 2, Appli
2	1478	93.9	330	2	US-08-525-864A-4	Sequence 4, Appli
3	776	49.3	469	3	US-08-753-007A-8	Sequence 8, Appli
4	776	49.3	469	3	US-09-398-496-8	Sequence 8, Appli
5	776	49.3	647	3	US-08-753-007A-32	Sequence 32, Appli
6	776	49.3	647	3	US-09-398-496-32	Sequence 32, Appli
7	736	46.8	407	3	US-08-753-007A-6	Sequence 6, Appli
8	736	46.8	407	3	US-09-398-496-6	Sequence 6, Appli
9	716	45.5	181	3	US-08-753-007A-4	Sequence 4, Appli
10	716	45.5	181	3	US-09-398-496-4	Sequence 4, Appli
11	716	45.5	605	3	US-08-753-007A-2	Sequence 2, Appli

12	716	45.5	605	3	US-09-398-496-2	Sequence 2, Appli
13	545	34.6	411	3	US-08-470-339-189	Sequence 189, App
14	545	34.6	422	4	US-08-467-602-324	Sequence 324, App
15	545	34.6	456	4	US-08-467-602-366	Sequence 366, App
16	545	34.6	601	3	US-08-470-335-233	Sequence 233, App
17	545	34.6	601	4	US-08-467-602-323	Sequence 323, App
18	545	34.6	610	3	US-08-470-335-236	Sequence 236, App
19	545	34.6	610	4	US-08-467-602-332	Sequence 332, App
20	545	34.6	635	4	US-08-467-602-365	Sequence 365, App
21	545	34.6	644	4	US-08-467-602-374	Sequence 374, App
22	545	34.6	818	3	US-08-470-335-234	Sequence 234, App
23	545	34.6	818	4	US-08-467-602-321	Sequence 321, App
24	545	34.6	827	3	US-08-470-335-237	Sequence 237, App
25	545	34.6	827	4	US-08-467-602-333	Sequence 333, App
26	545	34.6	852	4	US-08-467-602-363	Sequence 363, App
27	545	34.6	861	4	US-08-467-602-375	Sequence 375, App
28	545	34.6	865	3	US-08-470-335-235	Sequence 235, App
29	545	34.6	865	4	US-08-467-602-322	Sequence 322, App
30	545	34.6	874	3	US-08-470-335-238	Sequence 238, App
31	545	34.6	874	4	US-08-467-602-334	Sequence 334, App
32	545	34.6	899	4	US-08-467-602-364	Sequence 364, App
33	545	34.6	908	4	US-08-467-602-376	Sequence 376, App
34	544	34.6	422	1	US-08-036-555B-170	Sequence 170, App
35	544	34.6	422	1	US-08-469-569-170	Sequence 170, App
36	544	34.6	422	1	US-08-428-926-3	Sequence 3, Appli
37	544	34.6	422	1	US-08-249-322A-170	Sequence 170, App
38	544	34.6	422	1	US-08-428-927-3	Sequence 3, Appli
39	544	34.6	422	1	US-08-428-298-3	Sequence 3, Appli
40	544	34.6	422	1	US-08-339-517-3	Sequence 3, Appli
41	544	34.6	422	1	US-08-469-526A-170	Sequence 170, App
42	544	34.6	422	2	US-08-734-591A-170	Sequence 170, App
43	544	34.6	422	2	US-08-469-660-170	Sequence 170, App
44	544	34.6	422	3	US-08-341-018-72	Sequence 72, Appl
45	544	34.6	422	3	US-08-470-335-170	Sequence 170, App

ALIGNMENTS

RESULT 1

US-08-525-864A-2

; Sequence 2, Application US/08525864A

; Patent No. 5912326

; GENERAL INFORMATION:

; APPLICANT: Chang, Han

; TITLE OF INVENTION: Cerebellum-derived Growth Factors, and Uses

; TITLE OF INVENTION: Related thereto

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 28 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  AscII (text)
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/525,864A
;      FILING DATE:  8-SEP-1995
;      CLASSIFICATION:  530
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Kara, Catherine J.
;      REGISTRATION NUMBER:  41,106
;      REFERENCE/DOCKET NUMBER:  HUI-017
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;      INFORMATION FOR SEQ ID NO:  2:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  754 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-525-864A-2

```

```

Query Match          98.3%;  Score 1547;  DB 2;  Length 754;
Best Local Similarity 97.7%;  Pred. No. 4.7e-132;
Matches 291;  Conservative 4;  Mismatches 3;  Indels 0;  Gaps 0;

```

```

Qy      1 MRRDPAPGFSM LFGVSLACYS PSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        |||||||
Db      1 MRRDPAPGFSM LFGVSLACYS PSLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
        |||||||
Db      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy      121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
        |:| ||||:|||||
Db      121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy      181 DGKELNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLVNSVS 240
        |||||||
Db      181 DGKELNRSRDIRIKYGNGRKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 240

Qy      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCCQFAMVNFS 298
        |||||||
Db      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCCQFAMVNFS 298

```

RESULT 2

US-08-525-864A-4

```

; Sequence 4, Application US/08525864A
; Patent No. 5912326
; GENERAL INFORMATION:
; APPLICANT:  Chang, Han
; TITLE OF INVENTION:  Cerebellum-derived Growth Factors, and Uses
; TITLE OF INVENTION:  Related thereto
; NUMBER OF SEQUENCES:  18
; CORRESPONDENCE ADDRESS:

```

```

;   ADDRESSEE:  LAHIVE & COCKFIELD
;   STREET:  28 State Street
;   CITY:  Boston
;   STATE:  Massachusetts
;   COUNTRY:  USA
;   ZIP:  02109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  AscII (text)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/525,864A
;   FILING DATE:  8-SEP-1995
;   CLASSIFICATION:  530
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Kara, Catherine J.
;   REGISTRATION NUMBER:  41,106
;   REFERENCE/DOCKET NUMBER:  HUI-017
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (617)227-7400
;   TELEFAX:  (617)742-4214
;   INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  330 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-525-864A-4

```

```

Query Match          93.9%;  Score 1478;  DB 2;  Length 330;
Best Local Similarity 96.2%;  Pred. No. 2.8e-126;
Matches 278;  Conservative 5;  Mismatches 6;  Indels 0;  Gaps 0;

```

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
      |:| ||||:||||||||||||||||||||||||||||||||||||||||||
Db    121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKMSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
      ||||||||||||||||||||||||||||||||||||||||||||:|||||
Db    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGLHVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRC 289
      |||||||||||||||||||||||||||||||||||||:| |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289

```

```

RESULT 3
US-08-753-007A-8

```

```

; Sequence 8, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;   TITLE OF INVENTION: AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Fish & Richardson P.C.
;     STREET: 225 Franklin Street
;     CITY: Boston
;     STATE: MA
;     COUNTRY: US
;     ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/753,007A
;     FILING DATE: 19-NOV-1996
;     CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/699,591
;     FILING DATE: 19-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Fasse, J. Peter
;     REGISTRATION NUMBER: 32,983
;     REFERENCE/DOCKET NUMBER: 07334/022001
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-542-5070
;     TELEFAX: 617-542-8906
;     TELEX:
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 469 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: not relevant
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     FRAGMENT TYPE: internal
US-08-753-007A-8

```

```

Query Match          49.3%;  Score 776;  DB 3;  Length 469;
Best Local Similarity 97.3%;  Pred. No. 2.7e-62;
Matches 144;  Conservative 1;  Mismatches 3;  Indels 0;  Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

```

Qy 262 VNGGVCYYIEGINQLSCKCPVGYTGDRC 289
 ||||| : |||
 Db 151 VNGGVCYYIEGINQLSCKCPNGFFGQRC 178

RESULT 4

US-09-398-496-8

; Sequence 8, Application US/09398496
 ; Patent No. 6133423
 ; GENERAL INFORMATION:
 ; APPLICANT: Gearing, David P.
 ; APPLICANT: Busfield, Samantha J.
 ; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
 ; TITLE OF INVENTION: AND USES THEREFOR
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: US
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/398,496
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/753,007
 ; FILING DATE: 19-NOV-1996
 ; APPLICATION NUMBER: 08/699,591
 ; FILING DATE: 19-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fasse, J. Peter
 ; REGISTRATION NUMBER: 32,983
 ; REFERENCE/DOCKET NUMBER: 07334/022001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-542-5070
 ; TELEFAX: 617-542-8906
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 469 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FRAGMENT TYPE: internal

US-09-398-496-8

Query Match 49.3%; Score 776; DB 3; Length 469;
 Best Local Similarity 97.3%; Pred. No. 2.7e-62;
 Matches 144; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNGRKN 201
 |||
 Db 31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNGRKN 90
 Qy 202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
 |||
 Db 91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150
 Qy 262 VNGGVCYYIEGINQLSCKCPVGYTGDRC 289
 ||| : |||
 Db 151 VNGGVCYYIEGINQLSCKCPNGFFGQRC 178

RESULT 5

US-08-753-007A-32

; Sequence 32, Application US/08753007A

; Patent No. 6074841

; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; APPLICANT: Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES

; TITLE OF INVENTION: AND USES THEREFOR

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/753,007A

; FILING DATE: 19-NOV-1996

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/699,591

; FILING DATE: 19-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, J. Peter

; REGISTRATION NUMBER: 32,983

; REFERENCE/DOCKET NUMBER: 07334/022001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; TELEX:

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 647 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear


```

; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-398-496-32

```

Query Match 49.3%; Score 776; DB 3; Length 647;
Best Local Similarity 97.3%; Pred. No. 4.2e-62;
Matches 144; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy	142	ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKN	201
Db	31	ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKN	90
Qy	202	SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC	261
Db	91	SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC	150
Qy	262	VNGGVCYYIEGINQLSCKCPVGYTGDRC	289
		:	
Db	151	VNGGVCYYIEGINQLSCKCPNGFFGQRC	178

RESULT 7

US-08-753-007A-6

; Sequence 6, Application US/08753007A

; Patent No. 6074841

; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; APPLICANT: Busfield, Samantha J.

10 TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES

; TITLE OF INVENTION: AND USES THEREFOR

; NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

```

; STREET: 225 Franklin Street

```

CITY: Boston

STATE: MA

COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Diskette

```

```

;      MEDICAL FILE:  DISCLOSED
;      COMPUTER:  IBM Compatible

```

```

;      OPERATING SYSTEM:  DOS

```

```

;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0

```

```

; CURRENT APPLICATION DATA:

```

: APPLICATION NUMBER: US/08/753,007A

APPLICATION NUMBER: 9570
FILING DATE: 19-NOV-1996

: CLASSIFICATION: 536

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-753-007A-6

```

```

Query Match          46.8%; Score 736; DB 3; Length 407;
Best Local Similarity 97.1%; Pred. No. 9.7e-59;
Matches 136; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGN GRKNSRLQFNKV 209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGN GRKNSRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPVG YTGDR C 289
          |||||||||||| | : | ||
Db      121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 8

US-09-398-496-6

```

; Sequence 6, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,496
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/753,007
; FILING DATE: 19-NOV-1996
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-398-496-6

```

```

Query Match          46.8%; Score 736; DB 3; Length 407;
Best Local Similarity 97.1%; Pred. No. 9.7e-59;
Matches 136; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNGRKN SRLQFNKV 209
          |||
Db       1 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNGRKN SRLQFNKV 60

QY      210 KVEDAGEYVCEAENILGKDTVGRGRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          |||
Db       61 KVEDAGEYVCEAENILGKDTVGRGRLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

QY      270 IEGINQLSCKCPVGYTGDR 289
          |||
Db       121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 9

US-08-753-007A-4

```

; Sequence 4, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,007A
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-753-007A-4

```

```

Query Match          45.5%; Score 716; DB 3; Length 181;
Best Local Similarity 94.3%; Pred. No. 2.1e-57;
Matches 132; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db       1  MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db       61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPVGYTGDRC 289
          ||||| ||: ||
Db      121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 10

US-09-398-496-4

```

; Sequence 4, Application US/09398496
; Patent No. 6133423

```

```

; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,496
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/753,007
; FILING DATE: 19-NOV-1996
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-398-496-4

```

```

Query Match          45.5%; Score 716; DB 3; Length 181;
Best Local Similarity 94.3%; Pred. No. 2.1e-57;
Matches 132; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db      1 MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

```

Qy 270 IEGINQLSCKCPVGYTGDRC 289
||||||| | : | ||
Db 121 IEGINQLSCKCPNGFFGQRC 140

RESULT 11

US-08-753-007A-2

; Sequence 2, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,007A
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 605 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal

US-08-753-007A-2

Query Match 45.5%; Score 716; DB 3; Length 605;
Best Local Similarity 94.3%; Pred. No. 1.1e-56;
Matches 132; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKNRLQFNKV 209

```

Db          1 MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKNSRLQFNKV 60
Qy          210 KVEDAGEYVCEAENILGKDTVGRGLVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
              :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db          61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120
Qy          270 IEGINQLSCKCPVGYTGDR 289
              |||||:|:|
Db          121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 12

US-09-398-496-2

; Sequence 2, Application US/09398496

; Patent No. 6133423

; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; APPLICANT: Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES

; TITLE OF INVENTION: AND USES THEREFOR

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/398,496

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/753,007

; FILING DATE: 19-NOV-1996

; APPLICATION NUMBER: 08/699,591

; FILING DATE: 19-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, J. Peter

; REGISTRATION NUMBER: 32,983

; REFERENCE/DOCKET NUMBER: 07334/022001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 605 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear


```
; MOLECULE TYPE:  protein
; FRAGMENT TYPE:  internal
US-09-398-496-2
```

Query Match 45.5%; Score 716; DB 3; Length 605;
Best Local Similarity 94.3%; Pred. No. 1.1e-56;
Matches 132; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

Qy      150 MKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db      1   MKSQGTGEVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPVGYTGDR C 289
          |||||: |||
Db      121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 13

US-08-470-339-189

; Sequence 189, Application US/08470339C

; Patent No. 6232286

; GENERAL INFORMATION:

; APPLICANT: GOODEARL, ANDREW

; APPLICANT: STROOBANT, PAUL

; APPLICANT: MINGHETTI, LUISA

; APPLICANT: WATERFIELD, MICHAEL

; APPLICANT: MARCHIONNI, MARK

; APPLICANT: CHEN, MARIO S.

; APPLICANT: HILES, IAN

; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR

7. TITLE OF INVENTION: PREPARATION AND USE

; FILE REFERENCE: 04585/002008

: CURRENT APPLICATION NUMBER: US/08/470,339C

CURRENT FILING DATE: 1995-06-06

: EARLIER APPLICATION NUMBER: 08/036,555

; EARLIER FILING DATE: 1993-03-24

: EARLIER APPLICATION NUMBER: 07/940,389

: EARLIER FILING DATE: 1992-09-03

: EARLIER APPLICATION NUMBER: 07/907,138

```

: EARLIER FILING DATE: 1992-06-30

```

: EARLIER APPLICATION NUMBER: 07/863,703

EARLIER FILING DATE: 1992-04-03

```

; EARLIER FILING DATE: 1992 04 05
: EARLIER APPLICATION NUMBER: 91 07566.3 GB

```

```

; EARLIER APPLICATION NUMBER: 91
: EARLIER FILING DATE: 1999-04-10

```

```

;   EARLIER FILING DATE: 1999
:   NUMBER OF SEQ ID NOS: 226

```

```
; NUMBER OF SEQ ID NOS: 220
: SOFTWARE: FastSEO for Windows Version 4.0
```

; SEO ID NO 189

```
; SEQ ID NO 189
; LENGTH: 411
```

```

; LENGTH: 4
: TYPE: PRT

```

```
; ORGANISM: Homo sapiens
```

US-08-470-339-189

Query Match 34.6%; Score 545; DB 3; Length 411;

Best Local Similarity 35.6%; Pred. No. 2.2e-41;
Matches 127; Conservative 62; Mismatches 90; Indels 78; Gaps 13;

```

Qy      15  GVSLACYS--PSLKSVDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
      |  | : |||  ||: |||: | :| ||:||||  || :  | | :  ||
Db      58  GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116

Qy      60  PPASGRVA-----LVKVLDKWPLRSGLQ 83
      |||:| |  ||||  | ::|||:
Db     117  PPAAGPRALGPPAEEPLLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLK 176

Qy      84  REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
      :: ::|  || | : : ||||:|  : | | : :| ||:| | :|
Db     177  KDSLLTVRLGTWGHPAFPSCGRCLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235

Qy     129  LKKEVGKILCTDCATRPKLKKMKSQTGQVGEEKSLKCEAAAGNPQPSYRWFKDGKELNRS 188
      ||||| ::||  || | :||:||||  | | | :|| ::  ::|||:| ||||
Db     236  LKKEVSRVLCRKRCALPPRLKEMKSQESAAGSKVLVRCETSSEYSSSLRFKWFKNGNELNRK 295

Qy     189  ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
      ::|:|:  | : | | : | : | :| : : | | :  : :  : | :|
Db     296  NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNATSTS 353

Qy     246  WSG--HARKCNETAKSYCVNGGVVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF 297
      :|  |  || |  | :|||  | : :: ::  |  ||||  :|||  | :| :|
Db     354  TTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCQNYVMASF 410

```

RESULT 14

US-08-467-602-324

; Sequence 324, Application US/08467602C

; Patent No. 6444642

; GENERAL INFORMATION:

; APPLICANT: Sklar, Robert

; APPLICANT: Marchionni, Mark

; APPLICANT: Gwynne, David I.

; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND

; TITLE OF INVENTION: DISORDERS

; FILE REFERENCE: 04585/028003

; CURRENT APPLICATION NUMBER: US/08/467,602C

; CURRENT FILING DATE: 1995-06-06

; EARLIER APPLICATION NUMBER: 08/209,204

; EARLIER FILING DATE: 1994-03-08

; EARLIER APPLICATION NUMBER: 08/059,022

; EARLIER FILING DATE: 1993-05-06

; NUMBER OF SEQ ID NOS: 420

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 324

; LENGTH: 422

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-467-602-324

Query Match 34.6%; Score 545; DB 4; Length 422;

Best Local Similarity 35.6%; Pred. No. 2.3e-41;

Matches 127; Conservative 62; Mismatches 90; Indels 78; Gaps 13;

Qy 15 GVSLACYS--PSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
 | |: ||| |: |||: | :| |: ||| | | : | | : ||
 Db 58 GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116
 Qy 60 PPASGRVA-----LVKVLDKWPLRSGGLQ 83
 |||:| | ||| | :|||:
 Db 117 PPAAGPRALGPPAEPELLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHVQVWAVKAGGLK 176
 Qy 84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
 :: :|| | | : : ||||:| | : | | : | ||:| |:
 Db 177 KDSLLTVRLGTWGHAPFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235
 Qy 129 LKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRS 188
 ||||| :|| | | :|||:| | | | :|| :|||:| ||||
 Db 236 LKKEVSRVLCKRCALPPRLKEMKSQESAAGSKLVLCETSSEYSSSLRFKWKFNKNELNRK 295
 Qy 189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
 :|:|: | : | | : | :|||:| : | | : : : | :|
 Db 296 NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSANITIVESNATSTS 353
 Qy 246 WSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCQQFAMVNF 297
 :| | | | | :||| | : : : | | | | :||| | :| :|
 Db 354 TTGTSHLVKCAEKEKTFVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCONYVMASF 410

RESULT 15

US-08-467-602-366

; Sequence 366, Application US/08467602C

; Patent No. 6444642

; GENERAL INFORMATION:

; APPLICANT: Sklar, Robert

; APPLICANT: Marchionni, Mark

; APPLICANT: Gwynne, David I.

; TITLE OF INVENTION: METHODS FOR TREATING MUSCLE DISEASES AND

; TITLE OF INVENTION: DISORDERS

; FILE REFERENCE: 04585/028003

; CURRENT APPLICATION NUMBER: US/08/467,602C

; CURRENT FILING DATE: 1995-06-06

; EARLIER APPLICATION NUMBER: 08/209,204

; EARLIER FILING DATE: 1994-03-08

; EARLIER APPLICATION NUMBER: 08/059,022

; EARLIER FILING DATE: 1993-05-06

; NUMBER OF SEQ ID NOS: 420

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 366

; LENGTH: 456

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (34)...(34)

; OTHER INFORMATION: Xaa is any amino acid

US-08-467-602-366

Query Match 34.6%; Score 545; DB 4; Length 456;

Best Local Similarity 35.6%; Pred. No. 2.5e-41;

Matches 127; Conservative 62; Mismatches 90; Indels 78; Gaps 13;

Qy 15 GVSLACYS--PSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
 | |: ||| |: |||: | :| ||:|||| | | : | | : ||
 Db 92 GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 150
 Qy 60 PPASGRVA-----LVKVLDKWPLRSGGLQ 83
 |||:| | |||| | ::|||:
 Db 151 PPAAGPRALGPPAEPELLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHVQVWAVKAGGLK 210
 Qy 84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
 ::::| || |: : ||||:| | : | |: :| ||:| |:|
 Db 211 KDSLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 269
 Qy 129 LKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS 188
 |||| ::|| || |:||:|||| | | |:|| :: ::|||:| ||||
 Db 270 LKKEVSRVLCKRCALPPRLKEMKSQESAAGSKLVLCETSSSEYSSLRFKWFKNGNELNRK 329
 Qy 189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
 ::|:|: |: | |: || : |:|||:|: : || |: : : : | :|
 Db 330 NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNATSTS 387
 Qy 246 WSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF 297
 :| | || | ::|||| |: :: :: | |||| :||||| : | :|
 Db 388 TTGTSHLVKCAEKEKTFVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCCQNYVMASF 444

Search completed: August 17, 2004, 14:14:02
 Job time : 15.7102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:06:50 ; Search time 13.7611 Seconds
(without alignments)
2083.044 Million cell updates/sec

Title: US-09-864-675-4
Perfect score: 1574
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....KCPVGYTGDRCCQFAMVNFS 298

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1505	95.6	850	2	JC5700	ErbB kinase activa
2	1470	93.4	860	2	JC5702	ErbB kinase activa
3	1470	93.4	868	2	JC5701	ErbB kinase activa
4	544	34.6	422	2	S32357	glial growth facto
5	306.5	19.5	241	2	D43273	heregulin precurs
6	306.5	19.5	637	2	C43273	heregulin precurs
7	306.5	19.5	645	2	B43273	heregulin, splice
8	305.5	19.4	602	2	A45769	acetylcholine rece
9	299	19.0	230	2	A56210	neu differentiat
10	299	19.0	636	2	I61718	neu differentiat
11	299	19.0	662	2	I61722	neu differentiat
12	292	18.6	241	2	S32359	glial growth facto
13	280.5	17.8	462	2	I38404	neu differentiat

14	279.5	17.8	640	2	A43273	heregulin precurs
15	273	17.3	639	2	I61719	neu differentiatio
16	211.5	13.4	175	2	I38408	neu differentiatio
17	196.5	12.5	125	2	S62676	heregulin isoform
18	194	12.3	57	2	PC4415	ErbB kinase activa
19	164.5	10.5	5175	2	T20992	hypothetical prote
20	164.5	10.5	5198	2	T43290	hemicentin precurs
21	152	9.7	296	2	A56943	sensory/motor neur
22	151.5	9.6	6658	2	T13931	projectin - fruit
23	151	9.6	1323	2	PN0568	connectin 3B - chi
24	150	9.5	1217	1	EGMSMG	epidermal growth f
25	145.5	9.2	1896	2	T08851	Down syndrome cell
26	144	9.1	338	2	JC4776	limbic-system-asso
27	143.5	9.1	1612	2	T30805	dutt1 protein - mo
28	140.5	8.9	1651	2	T14160	transmembrane rece
29	139.5	8.9	7962	2	I38346	elastic titin - hu
30	136.5	8.7	919	2	T32541	unc-5 protein - Ca
31	136.5	8.7	947	1	B44294	unc-5 protein, lon
32	135.5	8.6	1356	2	JC1402	protein-tyrosine k
33	135	8.6	4391	2	A38096	perlecan precursor
34	133.5	8.5	1367	2	A41228	protein-tyrosine k
35	133.5	8.5	6642	2	T29757	protein UNC-89 - C
36	132.5	8.4	1133	1	EGRT	epidermal growth f
37	132.5	8.4	4162	2	T42633	connectin/titin -
38	131	8.3	125	2	I38405	neu differentiatio
39	131	8.3	338	2	JC1238	opioid-binding pro
40	131	8.3	345	2	JC1239	opioid-binding pro
41	130.5	8.3	940	2	A40985	projectin - fruit
42	129.5	8.2	1018	2	A54744	contactin 1 precur
43	129.5	8.2	1064	2	A40136	fibropellin Ia - s
44	129.5	8.2	2783	2	T34416	hypothetical prote
45	129	8.2	345	2	JC4025	opioid-binding cel

ALIGNMENTS

RESULT 1

JC5700

ErbB kinase activator alpha, brain and thymus - human

C;Species: Homo sapiens (man)

C;Date: 25-Nov-1997 #sequence_revision 25-Nov-1997 #text_change 08-Sep-2002

C;Accession: JC5700

R;Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyagawa, J.; Matsushita, N.; Nagatsu, T.; Taniguchi, N.; Ishiguro, H. J. Biochem. 122, 675-680, 1997

A;Title: A novel brain-derived member of the epidermal growth factor family that interacts with ErbB3 and ErbB4.

A;Reference number: JC5700; MUID:98006324; PMID:9348101

A;Accession: JC5700

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-850 <HIG>

A;Cross-references: DDBJ:AB005060; NID:g2626738; PIDN:BAA23417.1; PID:g2626739

A;Experimental source: SK-NSH cell

C;Comment: This protein is a member of the epidermal growth factor family. It is functionally similar to neurogulin in terms of directly activating ErbB4,

transactivating ErbB1, B2 and B3, and stimulating the differentiation of MDA-MB-453 cells.

C;Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology; immunoglobulin homology

C;Keywords: glycoprotein

F;258-311/Domain: Ig-like #status predicted <IGL>

F;345-381/Domain: EGF homology <EGF>

F;346-381/Domain: EGF-like #status predicted <EGF2>

F;147,278,451/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 95.6%; Score 1505; DB 2; Length 850;

Best Local Similarity 98.6%; Pred. No. 1.5e-107;

Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      |||
Db     93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
      |||
Db    153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 212

Qy    121 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
      |||
Db    213 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272

Qy    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
      |||
Db    273 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 332

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLSCKCPVGYTGDR 289
      |||
Db    333 TTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLSCKCPNGFFGQRC 381
```

RESULT 2

JC5702

ErbB kinase activator alpha2a, brain and thymus - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 25-Nov-1997 #sequence_revision 25-Nov-1997 #text_change 08-Sep-2002

C;Accession: JC5702; PC4417

R;Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyagawa, J.; Matsushita, N.; Nagatsu, T.; Taniguchi, N.; Ishiguro, H. J. Biochem. 122, 675-680, 1997

A;Title: A novel brain-derived member of the epidermal growth factor family that interacts with ErbB3 and ErbB4.

A;Reference number: JC5700; MUID:98006324; PMID:9348101

A;Accession: JC5702

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-860 <HI1>

A;Cross-references: DDBJ:D89996; NID:g2605631; PIDN:BAA23345.1; PID:g2605632

A;Experimental source: PC-12 cell

A;Accession: PC4417

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 'F',212-213,223-860 <HI2>

A;Residues: 128-162 <HI2>

A;Experimental source: PC-12 cell

C;Comment: This protein is a member of the epidermal growth factor family. It is functionally similar to neurogulin in terms of directly activating ErbB4, transactivating ErbB1, B2 and B3, and stimulating the differentiation of MDA-MB-453 cells.

C;Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology; immunoglobulin homology

F;361-397/Domain: EGF homology <EGF>

Query Match 93.4%; Score 1470; DB 2; Length 868;

Best Local Similarity 95.8%; Pred. No. 7.6e-105;

Matches 277; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
      |||
Db    109 MRRDPAPGSSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 168

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
      |||
Db    169 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAF 228

Qy    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWF 180
      |||
Db    229 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWF 288

Qy    181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
      |||
Db    289 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 348

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR 289
      |||
Db    349 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 397
```

RESULT 4

S32357

glial growth factor - human

C;Species: Homo sapiens (man)

C;Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 08-Sep-2002

C;Accession: S32357

R;Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.; Hendricks, M.; Danehy, F.; Misumi, D.; Sudhalter, J.; Kobayashi, K.; Wroblewski, D.; Lynch, C.; Baldassare, M.; Hiles, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield, M.D.; Stroobant, P.; Gwynne, D.

Nature 362, 312-318, 1993

A;Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.

A;Reference number: S32357; MUID:93205115; PMID:8096067

A;Accession: S32357

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-422 <MAR>

A;Cross-references: GB:L12260; NID:g292047; PIDN:AAB59622.1; PID:g292048

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

F;363-402/Domain: EGF homology <EGF>

Query Match 34.6%; Score 544; DB 2; Length 422;
 Best Local Similarity 35.6%; Pred. No. 3.4e-34;
 Matches 127; Conservative 62; Mismatches 90; Indels 78; Gaps 13;

```

Qy      15 GVSLACYS--PSLKSVDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
      | |: ||| ||: |||: | :| |: |||      | | :      | | :      ||
Db      58 GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116

Qy      60 PPASGRVA-----LVKVLDKWPLRSGGLQ 83
      |||:| |      |||      | ::|||:
Db     117 PPAAGPRALGPPAEPELLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLK 176

Qy      84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
      :: ::|      || |: : ||||:|      : | |: :| ||:| |:|
Db     177 KDSLLTVRLGTWGHFAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235

Qy     129 LKKEVGKILCTDCATRPKLKMKMSQTGQVGKQSLKCEAAAGNPQPSYRWFKDGKELNRS 188
      ||||| ::|| || |:|:||||      | | |:| :|      ::|||:| |||
Db     236 LKKEVSRVLCKRCALPPQLKEMKSQESAAGSKLVLCETSSEYSSLRFKWFKNGNELNRK 295

Qy     189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
      ::|:|: | : | |: || : |:|:|:| : || |: : : : | :|
Db     296 NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSASANITIVESNATSTS 353

Qy     246 WSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCQQFAMVNF 297
      :| | || | |:| ||| |: :: :| ||| : ||||| : | :|
Db     354 TTGTSHLVKCAEKEKTFVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCQNYVMASF 410
  
```

RESULT 5

D43273

heregulin precursor, splice form beta-3 - human

N;Alternate names: glial growth factor HRG-beta-3; neuregulin

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Sep-2002

C;Accession: D43273; S32358

R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.

Science 256, 1205-1210, 1992

A;Title: Identification of heregulin, a specific activator of p185(erbB2).

A;Reference number: A43273; MUID:92271253; PMID:1350381

A;Accession: D43273

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-241 <HOL>

R;Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.; Hendricks, M.; Danehy, F.; Misumi, D.; Sudhalter, J.; Kobayashi, K.; Wroblewski, D.; Lynch, C.; Baldassare, M.; Hiles, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield, M.D.; Stroobant, P.; Gwynne, D.

Nature 362, 312-318, 1993

A;Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.

A;Reference number: S32357; MUID:93205115; PMID:8096067

A;Accession: S32358
 A;Molecule type: mRNA
 A;Residues: 1-241 <MAR>
 A;Cross-references: GB:L12261; NID:g292049; PIDN:AAB59358.1; PID:g292050
 C;Genetics:
 A;Gene: GDB:HGL; GGF
 A;Cross-references: GDB:132656; OMIM:142445
 A;Map position: 8p22-8p11
 C;Superfamily: human heregulin; EGF homology; immunoglobulin homology
 C;Keywords: alternative splicing
 F;182-221/Domain: EGF homology <EGF>

Query Match 19.5%; Score 306.5; DB 2; Length 241;
 Best Local Similarity 33.0%; Pred. No. 2.9e-16;
 Matches 73; Conservative 36; Mismatches 61; Indels 51; Gaps 7;

Qy	126	GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW	178
		: : : : :	
Db	11	GKGKKKERGSGKKPESAAGSQSPALPPRLKEMKQESAAGSKLVLCETSSEYSSLRFKW	70
Qy	179	FKDGKELNRS---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAEINLGKDTVGRGL-	234
		: : : : : : : : :	
Db	71	FKNGNELNRKNKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANIT	128
Qy	235	-----YV-----NSVSTTLSSWSG--HARKCNETAKS	259
		: : :	
Db	129	IVESNEIITGMPASTEGAYVSSSPIRISVSTEGANTSSSTSTTGTSHLVKCAEKEKT	188
Qy	260	YCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF	297
		: : : : : : :	
Db	189	FCVNGGECFMVKDLSNPRLCKCPNEFTGDRCCQNYVMASF	229

RESULT 6

C43273

heregulin precursor, splice form beta-2 - human

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Sep-2002

C;Accession: C43273; I38407

R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.

Science 256, 1205-1210, 1992

A;Title: Identification of heregulin, a specific activator of p185(erbB2).

A;Reference number: A43273; MUID:92271253; PMID:1350381

A;Accession: C43273

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-637 <HOL>

R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S. Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I38407

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 119-406 <RES>
A;Cross-references: EMBL:U02329; NID:g408408; PIDN:AAA19954.1; PID:g408409
C;Genetics:
A;Gene: GDB:HGL
A;Cross-references: GDB:132656; OMIM:142445
A;Map position: 8p22-8p11
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology
C;Keywords: alternative splicing
F;182-221/Domain: EGF homology <EGF>

Query Match 19.5%; Score 306.5; DB 2; Length 637;
Best Local Similarity 33.0%; Pred. No. 9.1e-16;
Matches 73; Conservative 36; Mismatches 61; Indels 51; Gaps 7;

```

Qy      126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
      || ||| | : | |::||| | | |:: :: ::|
Db      11 GKGGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCESSSEYSSLRFKW 70

Qy      179 FKDGKELNRS---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      ||:| ||| ::::|: | |: || : |::|: : || |: :
Db      71 FKNGNELNRKNKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANIT 128

Qy      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || |::| :| :| | || | |:
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTTGTSHLVKCAEKEKT 188

Qy      260 YCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF 297
      :||| | : : : | ||| :||| | : | :|
Db      189 FCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCCQNYVMASF 229

```

RESULT 7

B43273

heregulin, splice form beta 1 - human

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Sep-2002

C;Accession: B43273; I38406

R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.

Science 256, 1205-1210, 1992

A;Title: Identification of heregulin, a specific activator of p185(erbB2).

A;Reference number: A43273; MUID:92271253; PMID:1350381

A;Accession: B43273

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-645 <HOL>

R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S. Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I38406

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'A',95-418,'F',420-645 <RES>
A;Cross-references: EMBL:U02328; NID:g408406; PIDN:AAA19953.1; PID:g408407
C;Genetics:
A;Gene: GDB:HGL
A;Cross-references: GDB:132656; OMIM:142445
A;Map position: 8p22-8p11
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology
C;Keywords: alternative splicing
F;182-221/Domain: EGF homology <EGF>

Query Match 19.5%; Score 306.5; DB 2; Length 645;
Best Local Similarity 33.0%; Pred. No. 9.2e-16;
Matches 73; Conservative 36; Mismatches 61; Indels 51; Gaps 7;

```

Qy      126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
      || ||| | : | |::||| | | |:: : :|
Db      11 GKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLCETSSEYSSLRFKW 70

Qy      179 FKDGKELNRS---RDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      ||:| ||| :::|: | : | | : |::|:| : | | : :
Db      71 FKNGNELNRKNKPQNIKIQKPKGK--SELRINKASLADSGEYMCKVISKLGNDSSASANIT 128

Qy      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || |::| :| :| | || | |:
Db      129 IVESNEIITGMPASTEGAYVSSSPIRISVSTEGANTSSSTSTTGTSHLVKCAEKEKT 188

Qy      260 YCVNGGVCYIIEGINQLS---CKCPVGYTGDRCQQFAMVNF 297
      :||| | : : : | ||| :||| | : | :|
Db      189 FCVNGGECFMVKDLNPSRYLCKCPNEFTGDRCONYVMASF 229

```

RESULT 8

A45769

acetylcholine receptor synthesis stimulator ARIA-1 precursor - chicken

C;Species: Gallus gallus (chicken)

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Sep-2002

C;Accession: A45769

R;Falls, D.L.; Rosen, K.M.; Corfas, G.; Lane, W.S.; Fischbach, G.D.

Cell 72, 801-815, 1993

A;Title: ARIA, a protein that stimulates acetylcholine receptor synthesis, is a member of the neu ligand family.

A;Reference number: A45769; MUID:93201602; PMID:8453670

A;Accession: A45769

A;Status: preliminary

A;Molecule type: mRNA; protein

A;Residues: 1-602 <FAL>

A;Cross-references: GB:L11264; NID:g212603; PIDN:AAA49037.1; PID:g212604

A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBIN:127787, NCBIP:127788)

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 19.4%; Score 305.5; DB 2; Length 602;
Best Local Similarity 33.0%; Pred. No. 1e-15;
Matches 65; Conservative 37; Mismatches 74; Indels 21; Gaps 5;

Qy 109 TEQPLVFKTAFAPLDTNGKNLKKEVVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAA 168
 :| || : | | | : ||||:|:| | |:| |:|
 Db 5 SEGPLQYSLAPTQTDVNS-----SYNTVPPKLKEMKNQEVAVGQKLVLRCETT 52

Qy 169 AGNPQPSYRWFKDGKEL---NRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENIL 225
 : | :| :|:|: | | :|:| :| | | : :|:| | :|
 Db 53 SEYPALRFKWLKNGKEITKKNRPENVKIP-KKQKKYSELHIYRATLADAGEYACRVSSKL 111

Qy 226 GKDTVGRGLYVNSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGI---NQLSCKC 280
 | |: : : : : | :| :| | ||: |:|:| | | | : : : :|:
 Db 112 GNDSTKASVIITDTNATSTSTTGTSHLTCKDIKQKAFVCVNGGECYMKVLDLPNPPRYLCRC 171

Qy 281 PVGYTGDRCCQFAMVNF 297
 | :|:| | | : | :|
 Db 172 PNEFTGDRCCQNYVMASF 188

RESULT 9

A56210

neu differentiation factor - rat (fragment)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 08-Sep-2002

C;Accession: A56210

R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S. Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: A56210

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-230 <RES>

A;Cross-references: EMBL:U02315; NID:g408380; PIDN:AAA19940.1; PID:g408381

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 19.0%; Score 299; DB 2; Length 230;
 Best Local Similarity 33.8%; Pred. No. 1e-15;
 Matches 67; Conservative 34; Mismatches 53; Indels 44; Gaps 6;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYNGN 198
 | |:|:| | | | | | |:| :| :|:|:| | | | :|:|: |
 Db 23 ALPPRLKEMKSQESAAGSKLVLCETSSEYSSLRFKWFKNGNELNRKNKPNIKIQKPG 82

Qy 199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
 : | |: | | : |:|:|:|: :| | : :|:|:| | | | :|
 Db 83 K--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSE 140

Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279
 |: |:| :| :| | | | |:|:| | | : : : :| | |
 Db 141 SPIRISVSTEGANTSSSTSTSTTGTSHLIKAEKEKTFVCVNGGECFTVKDLSNPSRYLCK 200

Qy 280 CPVGYTGDRCCQFAMVNF 297
 || :|:| | | : | :|
 Db 201 CPNEFTGDRCCQNYVMASF 218

RESULT 10

I61718

neu differentiation factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 08-Sep-2002

C;Accession: I61718; I61721; I61720

R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.;

Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.

Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I61718

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-636 <RES>

A;Cross-references: EMBL:U02318; NID:g408386; PIDN:AAA19943.1; PID:g408387

A;Accession: I61721

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-444,'A',446-636 <RE2>

A;Cross-references: EMBL:U02321; NID:g408392; PIDN:AAA19946.1; PID:g408393

A;Accession: I61720

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-298,386,'V',388,'TR',391 <RE3>

A;Cross-references: EMBL:U02320; NID:g408390; PIDN:AAA19945.1; PID:g408391

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

F;182-221/Domain: EGF homology <EGF>

Query Match 19.0%; Score 299; DB 2; Length 636;

Best Local Similarity 33.8%; Pred. No. 3.4e-15;

Matches 67; Conservative 34; Mismatches 53; Indels 44; Gaps 6;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS---RDIRIKYNG 198
| |:|:|:|:| | | |:|:| :: :|:|:| | | | :|:|: |

Db 34 ALPPRLKEMKSQESAAGSKVLRCETSSSEYSSLRFKWFKNGNELNRKNKPNIKIQKKPG 93

Qy 199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
: | |:|:| : |:|:|:|: : | | : : | |

Db 94 K--SELRINKASLADSGEYMCKVISKLGNDSSANITIVESNEFITGMPASTETAYVSSE 151

Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279
|: |:|:|:|:| | | | |:|:|:|:| |: :: :: | |

Db 152 SPIRISVSTEGANTSSSTSTSTGTSHLIKCAEKEKTFVCVNGGECFTVKDLSNPSRYLCK 211

Qy 280 CPVGYTGDRCCQFAMVNF 297
| | :|:|:|:|:| : | :|

Db 212 CPNEFTGDRCCQNYVMASF 229

RESULT 11

I61722

neu differentiation factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 08-Sep-2002

C;Accession: I61722

R;Wen, D.; Suggs, S.V.; Karunakaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.; Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S. Mol. Cell. Biol. 14, 1909-1919, 1994
A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.
A;Reference number: A56210; MUID:94158863; PMID:7509448
A;Accession: I61722
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-662 <RES>
A;Cross-references: EMBL:U02322; NID:g408394; PIDN:AAA19947.1; PID:g408395
C;Superfamily: human heregulin; EGF homology; immunoglobulin homology
F;182-221/Domain: EGF homology <EGF>

Query Match 19.0%; Score 299; DB 2; Length 662;
Best Local Similarity 33.8%; Pred. No. 3.6e-15;
Matches 67; Conservative 34; Mismatches 53; Indels 44; Gaps 6;

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYNG 198
      | |::|:|||| | | |::| :: :::|::| | || | :::|: |
Db      34 ALPPRLKEMKSQESAAGSKLVLRCEISSEYSSLRFKWKFKNGNELNRKNKPNIKIQKKPG 93

Qy      199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
      : | |::| : |::|::|: : || | : : ||
Db      94 K--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSE 151

Qy      237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVVCYYIEGINQLS---CK 279
      |: |::| :| :| | || | |::| || | |::: : : | ||
Db      152 SPIRISVSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVNGGECFTVKDLNPSRYLCK 211

Qy      280 CPVGYTGDRCCQFAMVNF 297
      || :||| | : | :|
Db      212 CPNEFTGDRCCQNYVMASF 229

```

RESULT 12

S32359

glial growth factor - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 19-Mar-1997 #sequence_revision 01-Aug-1997 #text_change 08-Sep-2002

C;Accession: S32359

R;Marchionni, M.A.; Goodearl, A.D.J.; Chen, M.S.; Bermingham-McDonogh, O.; Kirk, C.; Hendricks, M.; Danehy, F.; Misumi, D.; Sudhalter, J.; Kobayashi, K.; Wroblewski, D.; Lynch, C.; Baldassare, M.; Hiles, I.; Davis, J.B.; Hsuan, J.J.; Totty, N.F.; Otsu, M.; McBurney, R.N.; Waterfield, M.D.; Stroobant, P.; Gwynne, D.

Nature 362, 312-318, 1993

A;Title: Glial growth factors are alternatively spliced erbB2 ligands expressed in the nervous system.

A;Reference number: S32357; MUID:93205115; PMID:8096067

A;Accession: S32359

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-241 <MAR>

A;Cross-references: GB:L12259; NID:g289413; PIDN:AAA30540.1; PID:g289414

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

F;182-221/Domain: EGF homology <EGF>

Query Match 18.6%; Score 292; DB 2; Length 241;
Best Local Similarity 32.8%; Pred. No. 3.8e-15;
Matches 65; Conservative 37; Mismatches 52; Indels 44; Gaps 6;

```
Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKEL---NRSRDIRIKYGNG 198
      | | :|:|:| | | | | :|:| : : :|:|:| | | | : :|:|:| |
Db      34 ALPPRLKEMKSQESVAGSKLVLRCE TSSEYSSSLKFKWFKNGSELSRKNKPQNIKIQRPG 93

Qy      199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
      : | |: :| : |:|:|:|: : || |: : : :|:|:| |
Db      94 K--SELRISKASLADSGEYMCKVISKLGND SASANITIVESNEITTGMPASTETAYVSSE 151

Qy      237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCCYYIEGINQLS---CK 279
      |: |:| :| :| | | | | |:|:|:| | |: : : :| | |
Db      152 SPIRISVSTEGTNTSSSTSTSTAGTSHLVKCAEKEKTF CVNGGECFMVKDLSNPSRYLCK 211

Qy      280 CPVGYTGDRCCQFAMVNF 297
      || :|:|:| | : | :|
Db      212 CPNEFTGDRCCQNYVMASF 229
```

RESULT 13

I38404

neu differentiation factor - human

C;Species: Homo sapiens (man)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 08-Sep-2002

C;Accession: I38404

R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.;
Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.
Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu
differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I38404

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-462 <RES>

A;Cross-references: EMBL:U02326; NID:g408402; PIDN:AAA19951.1; PID:g408403

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 17.8%; Score 280.5; DB 2; Length 462;
Best Local Similarity 32.1%; Pred. No. 6.2e-14;
Matches 69; Conservative 35; Mismatches 60; Indels 51; Gaps 7;

```
Qy      126 GKNLKKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
      || ||| | : | | :|:|:| | | | | |:|:| : :|
Db      11 GKGGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCE TSSEYSSSLRFKW 70

Qy      179 FKDGKELNRS---RDIRIKYGNRKNNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      |:| | | | : |:|:| :| : | |: | | : |:|:|:|: : || |: :
Db      71 FKNGNELNRKNKPQNIKIQRKPGK--SELRINKASLADSGEYMCKVISKLGND SASANIT 128

Qy      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || | : |:| :| :| | | | | |:
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTSTGTSHLVKCAEKEKT 188
```

Qy 260 YCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQ 291
 :||||| |: :: :: | ||| |:|| || :
 Db 189 FCVNGGECFMVKDLNPNRYLCKCQPGFTGARCTE 223

RESULT 14

A43273

heregulin precursor, splice form alpha - human

N;Alternate names: breast cancer cell differentiation factor p45; Neu differentiation factor

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Sep-2002

C;Accession: A43273; A48498; A38155

R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansura, D.; Abadi, N.; Raab, H.; Lewis, G.D.; Shepard, H.M.; Kuang, W.J.; Wood, W.I.; Goeddel, D.V.; Vandlen, R.L.

Science 256, 1205-1210, 1992

A;Title: Identification of heregulin, a specific activator of p185(erbB2).

A;Reference number: A43273; MUID:92271253; PMID:1350381

A;Accession: A43273

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-640 <HOL>

A;Experimental source: breast tumor cell line, MDA-MB-231, ATCC HTB 26

A;Note: sequence extracted from NCBI backbone (NCBIP:103250)

R;Culouscou, J.M.; Plowman, G.D.; Carlton, G.W.; Green, J.M.; Shoyab, M. J. Biol. Chem. 268, 18407-18410, 1993

A;Title: Characterization of a breast cancer cell differentiation factor that specifically activates the HER4/p180(erbB4) receptor.

A;Reference number: A48498; MUID:93366731; PMID:7689552

A;Accession: A48498

A;Molecule type: protein

A;Residues: 20-21,'X',23-24,'XX',27-28 <CUL>

R;Peles, E.; Bacus, S.S.; Koski, R.A.; Lu, H.S.; Wen, D.; Ogden, S.G.; Levy, R.B.; Yarden, Y.

Cell 69, 205-216, 1992

A;Title: Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein that induces differentiation of mammary tumor cells.

A;Reference number: A38155; MUID:92208945; PMID:1348215

A;Accession: A38155

A;Molecule type: protein

A;Residues: 'X',15-16,'X',18-20,'RG',23-24,'GP',27,'E',29,'XP',32-36 <PEL>

A;Note: sequence extracted from NCBI backbone (NCBIP:91347)

C;Genetics:

A;Gene: GDB:HGL

A;Cross-references: GDB:132656; OMIM:142445

A;Map position: 8p22-8p11

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

C;Keywords: alternative splicing; glycoprotein

F;182-221/Domain: EGF homology <EGF>

Query Match 17.8%; Score 279.5; DB 2; Length 640;

Best Local Similarity 32.1%; Pred. No. 1.1e-13;

Matches 69; Conservative 35; Mismatches 60; Indels 51; Gaps 7;

Qy 126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178

```

      || ||| | : | :||:|||| | | |:| : :|
Db      11 GKGKKKERGSGKKPESAAGSQSPALPPQLKEMKSQESAAGSKLVLRCESTSEYSSLRFKW 70
QY      179 FKDGKELNRS---RDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
      ||:| |||| :||:| : | | : || : |:| |:| : || |: :
Db      71 FKNGNELNRKNKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDASANIT 128
QY      235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
      || | |:| :| :| | || | |:
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTTGTSHLVKCAEKEKT 188
QY      260 YCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQ 291
      :||||| |: : : : | ||| |:| || :
Db      189 FCVNGGECFMVKDLSNPSRYLCKCQPGFTGARCTE 223

```

RESULT 15

I61719

neu differentiation factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 08-Sep-2002

C;Accession: I61719; I61723; I61716; I61717; I61724; A38220

R;Wen, D.; Suggs, S.V.; Karunagaran, D.; Liu, N.; Cupples, R.L.; Luo, Y.;

Janssen, A.M.; Ben-Baruch, N.; Trollinger, D.B.; Jacobsen, V.L.; Meng, S.

Mol. Cell. Biol. 14, 1909-1919, 1994

A;Title: Structural and functional aspects of the multiplicity of Neu differentiation factors.

A;Reference number: A56210; MUID:94158863; PMID:7509448

A;Accession: I61719

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-639 <RES>

A;Cross-references: EMBL:U02319; NID:g408388; PIDN:AAA19944.1; PID:g408389

A;Accession: I61723

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-639 <RE2>

A;Cross-references: EMBL:U02323; NID:g408396; PIDN:AAA19948.1; PID:g408397

A;Accession: I61716

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-422, 'H', 'NL', 637-638, 'ELRRNKAYRSKCMQIQLSATHLRPSSITHLGFIL' <RE3>

A;Cross-references: EMBL:U02316; NID:g408382; PIDN:AAA19941.1; PID:g408383

A;Accession: I61717

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-422, 'H', 'NL', 637-638, 'ELRRNKAYRSKCMQIQLSATHLRPSSITHLGFIL' <RE4>

A;Cross-references: EMBL:U02317; NID:g408384; PIDN:AAA19942.1; PID:g408385

A;Accession: I61724

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-422 <RE5>

A;Cross-references: EMBL:U02324; NID:g408398; PIDN:AAA19949.1; PID:g408399

R;Wen, D.; Peles, E.; Cupples, R.; Suggs, S.V.; Bacus, S.S.; Luo, Y.; Trail, G.;

Hu, S.; Silbiger, S.M.; Levy, R.B.; Koski, R.A.; Lu, H.S.; Yarden, Y.

Cell 69, 559-572, 1992

A;Title: Neu differentiation factor: a transmembrane glycoprotein containing an EGF domain and an immunoglobulin homology unit.

A;Reference number: A38220; MUID:92257596; PMID:1349853

A;Accession: A38220

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-422 <WEN>

A;Note: sequence extracted from NCBI backbone (NCBIN:101767, NCBIP:101768)

C;Superfamily: human heregulin; EGF homology; immunoglobulin homology

Query Match 17.3%; Score 273; DB 2; Length 639;

Best Local Similarity 32.8%; Pred. No. 3.4e-13;

Matches 63; Conservative 33; Mismatches 52; Indels 44; Gaps 6;

```
Qy      142 ATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNG 198
      |  |:|:|:|  |  |  |:|  ::      ::|:|:|  |||  |:|:  |
Db      34 ALPPRLKEMKSQESAAGSKLVLRCESTSEYSSLRFKWFKNGNELNRKNKPENIKIQKKPG 93

Qy      199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV--- 236
      :  | |:  ||  : |:|:|:|:  : || |:  :      ||
Db      94 K--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSE 151

Qy      237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CK 279
      |: |:| :| :|  | || |  |:|:|:|  |: :: ::  |  ||
Db      152 SPIRISVSTEGANTSSSTSTTTGTSHLIKCAEKEKTFVCVNGGECFTVKDLNPSRYLCK 211

Qy      280 CPVGYTGDRCQQ 291
      |  |:|  ||  :
Db      212 CQPGFTGARCTE 223
```

Search completed: August 17, 2004, 14:13:20

Job time : 14.7611 secs

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:12:46 ; Search time 40.3344 Seconds
(without alignments)
2319.368 Million cell updates/sec

Title: US-09-864-675-4
Perfect score: 1574
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....KCPVGYTGDRCCQFAMVNF 298

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					
No.	Score	Match	Length	DB	ID	Description

1	1574	100.0	298	9	US-09-864-675-4	Sequence 4, Appli
2	1505	95.6	330	9	US-09-864-675-2	Sequence 2, Appli
3	1505	95.6	422	15	US-10-447-839A-3	Sequence 3, Appli
4	1505	95.6	426	15	US-10-447-839A-2	Sequence 2, Appli
5	1505	95.6	850	16	US-10-408-765A-610	Sequence 610, App
6	776	49.3	469	13	US-10-096-241-8	Sequence 8, Appli
7	776	49.3	647	13	US-10-096-241-32	Sequence 32, Appl
8	736	46.8	407	13	US-10-096-241-6	Sequence 6, Appli
9	716	45.5	181	13	US-10-096-241-4	Sequence 4, Appli
10	716	45.5	605	13	US-10-096-241-2	Sequence 2, Appli
11	544	34.6	422	8	US-08-736-019-170	Sequence 170, App
12	542	34.4	418	9	US-09-795-668-3	Sequence 3, Appli
13	542	34.4	418	9	US-09-795-686-3	Sequence 3, Appli
14	542	34.4	418	9	US-09-946-807-3	Sequence 3, Appli
15	504	32.0	139	13	US-10-096-241-33	Sequence 33, Appl
16	375	23.8	204	9	US-09-795-668-4	Sequence 4, Appli
17	375	23.8	204	9	US-09-795-686-4	Sequence 4, Appli
18	375	23.8	204	9	US-09-946-807-4	Sequence 4, Appli
19	325	20.6	501	14	US-10-290-578-10	Sequence 10, Appl
20	325	20.6	768	9	US-09-773-517-11	Sequence 11, Appl
21	325	20.6	768	9	US-09-792-025-11	Sequence 11, Appl
22	325	20.6	768	9	US-09-849-868-11	Sequence 11, Appl
23	325	20.6	768	9	US-09-808-602-85	Sequence 85, Appl
24	325	20.6	768	14	US-10-290-578-2	Sequence 2, Appli
25	325	20.6	768	14	US-10-453-183-11	Sequence 11, Appl
26	317	20.1	192	9	US-09-795-668-2	Sequence 2, Appli
27	317	20.1	192	9	US-09-795-686-2	Sequence 2, Appli
28	317	20.1	192	9	US-09-946-807-2	Sequence 2, Appli
29	306.5	19.5	239	9	US-09-795-668-18	Sequence 18, Appl
30	306.5	19.5	239	9	US-09-795-686-18	Sequence 18, Appl
31	306.5	19.5	239	9	US-09-946-807-18	Sequence 18, Appl
32	306.5	19.5	629	9	US-09-795-668-14	Sequence 14, Appl
33	306.5	19.5	629	9	US-09-795-686-14	Sequence 14, Appl
34	306.5	19.5	629	9	US-09-946-807-14	Sequence 14, Appl
35	306.5	19.5	637	9	US-09-795-668-13	Sequence 13, Appl
36	306.5	19.5	637	9	US-09-795-686-13	Sequence 13, Appl
37	306.5	19.5	637	9	US-09-946-807-13	Sequence 13, Appl
38	306.5	19.5	645	13	US-10-096-241-10	Sequence 10, Appl
39	305.5	19.4	241	9	US-09-773-517-7	Sequence 7, Appli
40	305.5	19.4	241	9	US-09-792-025-7	Sequence 7, Appli
41	305.5	19.4	241	9	US-09-849-868-7	Sequence 7, Appli
42	305.5	19.4	241	14	US-10-453-183-7	Sequence 7, Appli
43	305.5	19.4	420	9	US-09-773-517-9	Sequence 9, Appli
44	305.5	19.4	420	9	US-09-792-025-9	Sequence 9, Appli
45	305.5	19.4	420	9	US-09-849-868-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-864-675-4

; Sequence 4, Application US/09864675

; Patent No. US20020081286A1

; GENERAL INFORMATION:

; APPLICANT: Marchionni, Mark

```

; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
; FILE REFERENCE: 04585/049002
; CURRENT APPLICATION NUMBER: US/09/864,675
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/206,495
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-675-4

```

```

Query Match          100.0%; Score 1574; DB 9; Length 298;
Best Local Similarity 100.0%; Pred. No. 3.3e-121;
Matches 298; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
        |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
        |||
Db    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
        |||
Db    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR CQFAMVNFS 298
        |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR CQFAMVNFS 298

```

RESULT 2

US-09-864-675-2

```

; Sequence 2, Application US/09864675
; Patent No. US20020081286A1
; GENERAL INFORMATION:
; APPLICANT: Marchionni, Mark
; TITLE OF INVENTION: NRG-2 NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: POLYPEPTIDES, AND DIAGNOSTIC AND THERAPEUTIC METHODS
; FILE REFERENCE: 04585/049002
; CURRENT APPLICATION NUMBER: US/09/864,675
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/206,495
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 330

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-675-2

Query Match 95.6%; Score 1505; DB 9; Length 330;
Best Local Similarity 98.6%; Pred. No. 1.8e-115;
Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60

Qy     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
          |||
Db     61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120

Qy    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |||
Db    121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

Qy    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
          |||
Db    181 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240

Qy    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR 289
          |||
Db    241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 289
```

RESULT 3

US-10-447-839A-3

; Sequence 3, Application US/10447839A
; Publication No. US20040018181A1

; GENERAL INFORMATION:

; APPLICANT: Kufe, Donald W.

; APPLICANT: Kharbanda, Surender

; APPLICANT: Weitman, Steven D.

; TITLE OF INVENTION: MUC1 INTERFERENCE RNA COMPOSITIONS AND METHODS DERIVED THEREFROM

; FILE REFERENCE: 1000.05.009

; CURRENT APPLICATION NUMBER: US/10/447,839A

; CURRENT FILING DATE: 2003-05-29

; PRIOR APPLICATION NUMBER: 10/293,391

; PRIOR FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: 09/951,938

; PRIOR FILING DATE: 2001-09-11

; PRIOR APPLICATION NUMBER: 60/231,841

; PRIOR FILING DATE: 2000-09-11

; NUMBER OF SEQ ID NOS: 109

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 3

; LENGTH: 422

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-447-839A-3

Query Match 95.6%; Score 1505; DB 15; Length 422;

Best Local Similarity 98.6%; Pred. No. 2.4e-115;
Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
Db      93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
          |||
Db     153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212

Qy     121 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
          |||
Db     213 PLDTNGKNLKKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272

Qy     181 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVS 240
          |||
Db     273 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVS 332

Qy     241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRC 289
          |||
Db     333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 381
```

RESULT 4

US-10-447-839A-2

; Sequence 2, Application US/10447839A

; Publication No. US20040018181A1

; GENERAL INFORMATION:

; APPLICANT: Kufe, Donald W.

; APPLICANT: Kharbanda, Surrender

; APPLICANT: Weitman, Steven D.

; TITLE OF INVENTION: MUC1 INTERFERENCE RNA COMPOSITIONS AND METHODS DERIVED THEREFROM

; FILE REFERENCE: 1000.05.009

; CURRENT APPLICATION NUMBER: US/10/447,839A

; CURRENT FILING DATE: 2003-05-29

; PRIOR APPLICATION NUMBER: 10/293,391

; PRIOR FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: 09/951,938

; PRIOR FILING DATE: 2001-09-11

; PRIOR APPLICATION NUMBER: 60/231,841

; PRIOR FILING DATE: 2000-09-11

; NUMBER OF SEQ ID NOS: 109

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 426

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-447-839A-2

Query Match 95.6%; Score 1505; DB 15; Length 426;

Best Local Similarity 98.6%; Pred. No. 2.4e-115;

Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
          |||
```

Db 93 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

QY 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
 |||

Db 153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212

QY 121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
 |||

Db 213 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 272

QY 181 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
 |||

Db 273 DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 332

QY 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR 289
 ||| : ||

Db 333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 381

RESULT 5

US-10-408-765A-610

; Sequence 610, Application US/10408765A

; Publication No. US20040101874A1

; GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Fahy, Eoin D.

; APPLICANT: Zhang, Bing

; APPLICANT: Gibson, Bradford W.

; APPLICANT: Taylor, Steven W.

; APPLICANT: Glenn, Gary M.

; APPLICANT: Warnock, Dale E.

; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408,765A

; CURRENT FILING DATE: 2003-04-04

; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 610

; LENGTH: 850

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-408-765A-610

Query Match 95.6%; Score 1505; DB 16; Length 850;

Best Local Similarity 98.6%; Pred. No. 5.8e-115;

Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
 |||

Db 93 MRRDPAPGFSMLLFGVSLACYSPSLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152

QY 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
 |||

Db 153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 212

QY 121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180

Db	213	PLDTNGKNLKKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAAAGNPQPSYRWFK	272
Qy	181	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS	240
Db	273	DGKELNRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS	332
Qy	241	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDR	289
Db	333	TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC	381

RESULT 6

US-10-096-241-8

; Sequence 8, Application US/10096241

; Publication No. US20020127594A1

; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
AND USES THEREFOR

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/096,241

; FILING DATE: 12-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/699,591

; FILING DATE: 19-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, J. Peter

; REGISTRATION NUMBER: 32,983

; REFERENCE/DOCKET NUMBER: 07334/022001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 469 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-096-241-8

Query Match 49.3%; Score 776; DB 13; Length 469;
Best Local Similarity 97.3%; Pred. No. 2.4e-55;
Matches 144; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN 201
          |||
Db      31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          |||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPVGYTGDRG 289
          |||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRC 178
```

RESULT 7

US-10-096-241-32

; Sequence 32, Application US/10096241

; Publication No. US20020127594A1

; GENERAL INFORMATION:

; APPLICANT: Gearing, David P.

; Busfield, Samantha J.

; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
AND USES THEREFOR

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: US

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/096,241

; FILING DATE: 12-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/699,591

; FILING DATE: 19-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, J. Peter

; REGISTRATION NUMBER: 32,983

; REFERENCE/DOCKET NUMBER: 07334/022001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-5070

; TELEFAX: 617-542-8906

; TELEX: <Unknown>

```

; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 647 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     FRAGMENT TYPE: internal
;     SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-096-241-32

```

```

Query Match          49.3%; Score 776; DB 13; Length 647;
Best Local Similarity 97.3%; Pred. No. 3.6e-55;
Matches 144; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN 90

Qy      202 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 SRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYC 150

Qy      262 VNGGVCYYIEGINQLSCKCPVGYTGDRG 289
          ||||||||||||||||| | : | ||
Db      151 VNGGVCYYIEGINQLSCKCPNGFFGQRC 178

```

RESULT 8

US-10-096-241-6

```

; Sequence 6, Application US/10096241
; Publication No. US20020127594A1
; GENERAL INFORMATION:
;     APPLICANT: Gearing, David P.
;               Busfield, Samantha J.
;     TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;                       AND USES THEREFOR
;     NUMBER OF SEQUENCES: 33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Fish & Richardson P.C.
;       STREET: 225 Franklin Street
;       CITY: Boston
;       STATE: MA
;       COUNTRY: US
;       ZIP: 02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette
;       COMPUTER: IBM Compatible
;       OPERATING SYSTEM: DOS
;       SOFTWARE: FastSEQ Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/10/096,241
;       FILING DATE: 12-Mar-2002
;       CLASSIFICATION: <Unknown>
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/699,591
;       FILING DATE: 19-AUG-1996

```

```

;      ATTORNEY/AGENT INFORMATION:
;      NAME: Fasse, J. Peter
;      REGISTRATION NUMBER: 32,983
;      REFERENCE/DOCKET NUMBER: 07334/022001
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 617-542-5070
;      TELEFAX: 617-542-8906
;      TELEX: <Unknown>
;      INFORMATION FOR SEQ ID NO: 6:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 407 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      FRAGMENT TYPE: internal
;      SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-096-241-6

```

```

Query Match          46.8%;  Score 736;  DB 13;  Length 407;
Best Local Similarity 97.1%;  Pred. No. 3.9e-52;
Matches 136;  Conservative 1;  Mismatches 3;  Indels 0;  Gaps 0;

```

```

QY      150 MKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  MKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY      270 IEGINQLSCKCPVGYTGDRG 289
          ||||||||| | : | ||
Db      121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 9

US-10-096-241-4

```

; Sequence 4, Application US/10096241
; Publication No. US20020127594A1
;      GENERAL INFORMATION:
;      APPLICANT: Gearing, David P.
;      Busfield, Samantha J.
;      TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;      AND USES THEREFOR
;      NUMBER OF SEQUENCES: 33
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Fish & Richardson P.C.
;      STREET: 225 Franklin Street
;      CITY: Boston
;      STATE: MA
;      COUNTRY: US
;      ZIP: 02110-2804
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: DOS

```

```

;      SOFTWARE: FastSEQ Version 2.0
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/10/096,241
;      FILING DATE: 12-Mar-2002
;      CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 08/699,591
;      FILING DATE: 19-AUG-1996
;      ATTORNEY/AGENT INFORMATION:
;      NAME: Fasse, J. Peter
;      REGISTRATION NUMBER: 32,983
;      REFERENCE/DOCKET NUMBER: 07334/022001
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 617-542-5070
;      TELEFAX: 617-542-8906
;      TELEX: <Unknown>
;      INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 181 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      FRAGMENT TYPE: internal
;      SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-096-241-4

```

```

Query Match          45.5%;  Score 716;  DB 13;  Length 181;
Best Local Similarity 94.3%;  Pred. No. 6.2e-51;
Matches 132;  Conservative 4;  Mismatches 4;  Indels 0;  Gaps 0;

```

```

Qy      150 MKSQTGGVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRLQFNKV 209
          |||||:|||||
Db      1 MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKN SRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
          :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRGLHVN SVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPVGYTGDR C 289
          ||||| ||: ||
Db      121 IEGINQLSCKCPNGFFGQRC 140

```

RESULT 10

US-10-096-241-2

```

; Sequence 2, Application US/10096241
; Publication No. US20020127594A1
;      GENERAL INFORMATION:
;      APPLICANT: Gearing, David P.
;      Busfield, Samantha J.
;      TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;      AND USES THEREFOR
;      NUMBER OF SEQUENCES: 33
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Fish & Richardson P.C.
;      STREET: 225 Franklin Street

```

```

;          CITY: Boston
;          STATE: MA
;          COUNTRY: US
;          ZIP: 02110-2804
;    COMPUTER READABLE FORM:
;          MEDIUM TYPE: Diskette
;          COMPUTER: IBM Compatible
;          OPERATING SYSTEM: DOS
;          SOFTWARE: FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/10/096,241
;          FILING DATE: 12-Mar-2002
;          CLASSIFICATION: <Unknown>
;    PRIOR APPLICATION DATA:
;          APPLICATION NUMBER: 08/699,591
;          FILING DATE: 19-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;          NAME: Fasse, J. Peter
;          REGISTRATION NUMBER: 32,983
;          REFERENCE/DOCKET NUMBER: 07334/022001
;    TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 617-542-5070
;          TELEFAX: 617-542-8906
;          TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 2:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH: 605 amino acids
;        TYPE: amino acid
;        STRANDEDNESS: not relevant
;        TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      FRAGMENT TYPE: internal
;      SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-096-241-2

```

```

Query Match          45.5%;  Score 716;  DB 13;  Length 605;
Best Local Similarity 94.3%;  Pred. No. 2.8e-50;
Matches 132;  Conservative 4;  Mismatches 4;  Indels 0;  Gaps 0;

```

```

Qy      150 MKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNGRKNSRLQFNKV 209
        |||||:|||||
Db      1   MKSQTGEVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNVRKNSRLQFNKV 60

Qy      210 KVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 269
        :|||||:|||||
Db      61 RVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYY 120

Qy      270 IEGINQLSCKCPVGYTGDRG 289
        |||||:|
Db      121 IEGINQLSCKCPNGFFGQRC 140

```

```

RESULT 11
US-08-736-019-170
; Sequence 170, Application US/08736019
; Publication No. US20030207799A1
; GENERAL INFORMATION:

```


; APPLICANT: Goodearl, Andrew
; APPLICANT: Stroobant, Paul
; APPLICANT: Minghetti, Luisa
; APPLICANT: Waterfield, Michael
; APPLICANT: Marchionni, Mark
; APPLICANT: Chen, Mario
; APPLICANT: Hiles, Ian
; TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 189
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible Pentium
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/736,019
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/471,833
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 91 07566.3
; FILING DATE: 10-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER: 39,109
; REFERENCE/DOCKET NUMBER: 04585/00200Q
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 428-0200
; TELEFAX: (617) 428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-736-019-170

Query Match 34.6%; Score 544; DB 8; Length 422;
Best Local Similarity 35.6%; Pred. No. 2.5e-36;
Matches 127; Conservative 62; Mismatches 90; Indels 78; Gaps 13;

Qy 15 GVSLACYS--PSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
| |: ||| |: |||: | :| ||:|||| | | : | | : | |
Db 58 GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116

Qy 60 PPASGRVA-----LVKVLDDKWPLRSGGLQ 83
|||:| | ||| | :|||:
Db 117 PPAAGPRALGPPAEELLANGTVPSWPTAPVPSAGEPGEEAPYLVKVHVQVWAVKAGGLK 176

Qy 84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
:: :|| | | : : ||||:| | : | | : | ||: | :|
Db 177 KDSLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235

Qy 129 LKKEVGKILCTDCATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGGKELNRS 188
|||| :|| | | :||:|||| | | :|| : : |||:| |||
Db 236 LKKEVSRVLCRCALPPQLKEMKSQESAAGSKLVLCETSSEYSSLRFKWFKNGNELNRK 295

Qy 189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
:|:|: | : | | : | :|||:| : | | : : : | :|
Db 296 NKPQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNATSTS 353

Qy 246 WSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF 297
:| | || | | :||| | : : : | ||| :||| | :| :|
Db 354 TTGTSHLVKCAEKEKTFVNGGECFMVKDLSNPRLCKCPNEFTGDRCCQNYVMASF 410

RESULT 12

US-09-795-668-3

; Sequence 3, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-668-3

Query Match 34.4%; Score 542; DB 9; Length 418;
 Best Local Similarity 35.2%; Pred. No. 3.6e-36;
 Matches 125; Conservative 62; Mismatches 92; Indels 76; Gaps 12;

```

Qy      15 GVSLACYSPSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTREPP 61
      | |: | |: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
Db      56 GASV-CSPPSVGSGVQELAQRAAVVIEGKVHPQRRQOGALDRKAAAAAGEAGAWGGDREPP 114

Qy      62 ASGRVA-----LVKVLDKWPLRSGGLQRE 85
      |:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     115 AAGPRALGPPAEELLAAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLKKD 174

Qy      86 QVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKNLK 130
      :::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     175 SLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRNLK 233

Qy     131 KEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS-- 188
      ||| :|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     234 KEVSRVLCKRCALPPRLKEMKSQESAAGSKVLRLCETSSSEYSSLRFKWFKNGNELNRKNK 293

Qy     189 -RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWS 247
      ::|:|: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     294 PQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSANITIVESNATSTSTT 351

Qy     248 G--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF 297
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     352 GTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCKCPNEFTGDRCCQNYVMASF 406
  
```

RESULT 13

US-09-795-686-3

; Sequence 3, Application US/09795686

; Patent No. US20020094954A1

; GENERAL INFORMATION:

; APPLICANT: Stefansson, Hreinn

; APPLICANT: Steinthorsdottir, Valgerdur

; APPLICANT: Gulcher, Jeffrey R.

; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

; FILE REFERENCE: 2345.2005-001

; CURRENT APPLICATION NUMBER: US/09/795,686

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 09/515,715

; PRIOR FILING DATE: 2000-02-28

; NUMBER OF SEQ ID NOS: 1531

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 418

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-795-686-3

Query Match 34.4%; Score 542; DB 9; Length 418;

Best Local Similarity 35.2%; Pred. No. 3.6e-36;

Matches 125; Conservative 62; Mismatches 92; Indels 76; Gaps 12;

```

Qy      15 GVSLACYSPSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTREPP 61
      | |: | |: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
  
```

```

Db          56 GASV-CSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDREPP 114
Qy          62 ASGRVA-----LVKVLDDKWPLRSGGLQRE 85
           |:| |
Db          115 AAGPRALGPPAEPLLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLKKD 174
           ||| | :|||:
Qy          86 QVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKNLK 130
           :|| | | | : | | | | : | | | | |
Db          175 SLLTVRLGTWGHAPAFPSGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPLET-GRNLK 233
           || | : | | | : | | | | |
Qy          131 KEVGKILCTDCATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGKELNRS-- 188
           ||| :|| | | | :|||: | | | :| :| :| :|
Db          234 KEVSRVLCKRCALPPRLKEMKSQESAAGSKVLVRCETSSEYSSLRFKWFKNGNELNRKNK 293
           ||| :|| | | | :|||: | | | :| :| :| :|
Qy          189 -RDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWS 247
           :|||: | : | | : | : | : | : | : | : | : | : | : |
Db          294 PQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANITIVESNATSTSTT 351
           || | | | | :|||: | : | : | : | : | : | : | : |
Qy          248 G--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYTGDRCCQFAMVNF 297
           | | | | | :|||: | : | : | : | : | : | : | : |
Db          352 GTSHLVKCAEKEKTFVNGGECFMVKDLNPSRYLCKCPNEFTGDRCCQNYVMASF 406
           || | | | | :|||: | : | : | : | : | : | : | : |

```

RESULT 14

US-09-946-807-3

```

; Sequence 3, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-807-3

```

```

Query Match          34.4%; Score 542; DB 9; Length 418;
Best Local Similarity 35.2%; Pred. No. 3.6e-36;
Matches 125; Conservative 62; Mismatches 92; Indels 76; Gaps 12;

```

```

Qy          15 GVSLACYSPLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTREPP 61
           | | : | | | : | : | | : | | : | | : | |
Db          56 GASV-CSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDREPP 114
           ||| | :|||: | | | :| :| :| :|
Qy          62 ASGRVA-----LVKVLDDKWPLRSGGLQRE 85
           |:| |
           ||| | :|||:

```

Db	115	AAGPRALGPPAEELLAANGTVPSWPTAPVPSAGEPGEEAPYLVKVHQVWAVKAGGLKKD	174
Qy	86	QVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKNLK	130
		::: : : : :	
Db	175	SLLTVRLGTWGHPAFPSCGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRNLK	233
Qy	131	KEVGKILCTDCATRPKLKKMKSQTGQVGEEKQSLKCEAAAGNPQPSYRWFKDGKELNRS--	188
		:: : : : :: :: :	
Db	234	KEVSRVLCKRCALPPRLKEMKSQESAAGSKLVLRCESTSSEYSSLRFKWFKNGNELNRKNK	293
Qy	189	-RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWS	247
		::: : : : : : : : : : : : :	
Db	294	PQNIKIQQKPGK--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNATSTSTT	351
Qy	248	G--HARKCNETAKSYCVNGGVVCYYIEGINQLS--CKCPVGYTGDRCCQFAMVNF	297
		:: : : : : : : :	
Db	352	GTSHLVKCAEKEKTFVCVNGGECFMVKDLNPSRYLCKCPNEFTGDRCONYVMASF	406

RESULT 15

```

; Sequence 33, Application US/10096241
; Publication No. US20020127594A1
;   GENERAL INFORMATION:
;       APPLICANT: Gearing, David P.
;                   Busfield, Samantha J.
;   TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;                       AND USES THEREFOR
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Fish & Richardson P.C.
;       STREET: 225 Franklin Street
;       CITY: Boston
;       STATE: MA
;       COUNTRY: US
;       ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette
;       COMPUTER: IBM Compatible
;       OPERATING SYSTEM: DOS
;       SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/10/096,241
;       FILING DATE: 12-Mar-2002
;       CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/699,591
;       FILING DATE: 19-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Fasse, J. Peter
;       REGISTRATION NUMBER: 32,983
;       REFERENCE/DOCKET NUMBER: 07334/022001
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 617-542-5070
;       TELEFAX: 617-542-8906
;       TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 33:

```

```

;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 139 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-10-096-241-33

```

```

Query Match          32.0%;  Score 504;  DB 13;  Length 139;
Best Local Similarity 93.9%;  Pred. No. 1.2e-33;
Matches    92;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

```

```

Qy      192 RIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSGHAR 251
          ||||||||||||||||:||||||||||||||||||:||||||||||||||
Db       1 RIKYGNRKNRLQFNKVRVEDAGEYVCEAENILGKDTVGRGLHVNSVSTTLSSWSGHAR 60

Qy      252 KCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRG 289
          |||||||||||||||||||||:| ||
Db       61 KCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 98

```

```

Search completed: August 17, 2004, 14:22:30
Job time : 41.3344 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 14:05:35 ; Search time 33.6911 Seconds
(without alignments)
2790.781 Million cell updates/sec

Title: US-09-864-675-4
Perfect score: 1574
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....KCPVGYTGDRCQQFAMVNFS 298

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%
Result Query
No. Score Match Length DB ID Description

1	506	32.1	782	11	Q9ESA5	Q9esa5 rattus norv
2	476	30.2	342	11	Q9ESA1	Q9esa1 rattus norv
3	473	30.1	323	11	Q9ESA2	Q9esa2 rattus norv
4	471	29.9	317	11	Q9ESA3	Q9esa3 rattus norv
5	404.5	25.7	348	4	Q8NFN3	Q8nfn3 homo sapien
6	292	18.6	241	6	Q07112	Q07112 bos taurus
7	277	17.6	461	11	O35947	O35947 mesocricetu
8	269	17.1	54	11	Q810X1	Q810x1 mus musculu
9	234	14.9	211	11	Q8BKI8	Q8bki8 mus musculu
10	221	14.0	244	11	Q9ESA4	Q9esa4 rattus norv
11	200	12.7	79	11	Q810X2	Q810x2 mus musculu
12	180.5	11.5	167	4	Q8NFN2	Q8nfn2 homo sapien
13	166	10.5	8625	5	Q86GD6	Q86gd6 procambaru
14	164.5	10.5	5175	5	Q8I0L3	Q8i0l3 caenorhabdi
15	164.5	10.5	5198	5	O76518	O76518 caenorhabdi
16	158	10.0	1106	4	Q8WX93	Q8wx93 homo sapien
17	153.5	9.8	8943	5	Q9V4F7	Q9v4f7 drosophila
18	152	9.7	296	4	Q96IB3	Q96ib3 homo sapien
19	151.5	9.6	111	11	Q9ESA8	Q9esa8 rattus norv
20	151.5	9.6	6658	5	O76281	O76281 drosophila
21	151	9.6	298	11	Q9JI59	Q9ji59 mus musculu
22	151	9.6	298	11	Q8C5K9	Q8c5k9 mus musculu
23	151	9.6	410	4	Q8N1M2	Q8n1m2 homo sapien
24	151	9.6	1323	13	Q08476	Q08476 gallus gall
25	150	9.5	76	11	Q810X0	Q810x0 mus musculu
26	150	9.5	256	11	Q9ESA6	Q9esa6 rattus norv
27	150	9.5	296	11	Q8BX76	Q8bx76 mus musculu
28	150	9.5	298	11	Q8CE95	Q8ce95 mus musculu
29	150	9.5	700	11	Q9ESB1	Q9esb1 rattus norv
30	150	9.5	1200	11	Q8VD07	Q8vd07 mus musculu
31	149	9.5	136	11	Q9ESA7	Q9esa7 rattus norv
32	146.5	9.3	330	13	Q90Z42	Q90z42 gallus gall
33	144.5	9.2	754	11	Q8BZ76	Q8bz76 mus musculu
34	144	9.1	338	4	Q8IV49	Q8iv49 homo sapien
35	143.5	9.1	507	4	Q96K90	Q96k90 homo sapien
36	143.5	9.1	1320	4	Q96KF5	Q96kf5 homo sapien
37	143.5	9.1	1320	4	Q86TC9	Q86tc9 homo sapien
38	143.5	9.1	1391	4	Q8N3L4	Q8n3l4 homo sapien
39	143.5	9.1	1612	11	O89026	O89026 mus musculu
40	143.5	9.1	1651	4	Q9Y6N7	Q9y6n7 homo sapien
41	140.5	8.9	1651	11	O55005	O55005 rattus norv
42	140.5	8.9	2013	11	Q8VHZ8	Q8vzh8 rattus norv
43	140.5	8.9	2013	11	Q9ERC8	Q9erc8 mus musculu
44	140.5	8.9	3950	6	Q7YRF5	Q7yrf5 canis famil
45	140	8.9	341	11	Q8BLK3	Q8blk3 mus musculu

ALIGNMENTS

RESULT 1

Q9ESA5

ID Q9ESA5 PRELIMINARY; PRT; 782 AA.
AC Q9ESA5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Glial growth factor beta 1a (Fragment).
 GN NRG1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord, and Brain stem;
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
 RA Frohnert P.W.;
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF194993; AAG28433.1; -.
 DR HSSP; Q12784; 1HRE.
 DR GO; GO:0005102; F:receptor binding; IEA.
 DR GO; GO:0005351; F:sugar porter activity; IEA.
 DR GO; GO:0009790; P:embryonic development; IEA.
 DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR002114; HPr_SerP_S.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002154; Neuregulin.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF02158; Neuregulin; 1.
 DR PRINTS; PR01089; NEUREGULIN.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 DR PROSITE; PS00589; PTS_HPR_SER; 1.
 KW EGF-like domain; Immunoglobulin domain.
 FT NON_TER 1 1
 SQ SEQUENCE 782 AA; 86036 MW; F6174A68F4E27BDE CRC64;

Query Match 32.1%; Score 506; DB 11; Length 782;
 Best Local Similarity 32.5%; Pred. No. 1.9e-36;
 Matches 121; Conservative 60; Mismatches 89; Indels 102; Gaps 12;

Qy 23 PSLKSVQDQAYKAPVVVEGKV-----QG----- 45
 ||: |||: | :| ||:|||| ||
 Db 1 PSVGSVQELARRAAVVIEGKVHPPRRQQGALDRKAAGEAGAGARDQPVQDSPPSQDPLPA 60
 Qy 46 ---LVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92
 :| || : : | ||| | :|||: : :| |
 Db 61 VNWTLP TGGPEPST--DQPGDPAPYLKVKVQVWAVKAGGLKKDSLTLVRLDTWGHPAFPS 118
 Qy 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKL 147
 | | : | |||: || | : :| ||: | :||| || :| | :|
 Db 119 CGRLKEDSR YIFFMEPDANSSGRAPPAFRASFPLET-GRNLKKEVSRVLCRKRCALPPRL 177
 Qy 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNRKNRSRL 204
 |:|||| | | |:| : : :|||: | ||| :|:| :| :|

```

Db      178 KEMKSQESAAGSKLVLRCE TSSEYSSLRFKWFKNGNELNRKNKPENIKIQKPGK--SEL 235
Qy      205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236
      : || : |::|:|: : || |: : ||
Db      236 RINKASLADSGEYMCKVISKLGND SASANITIVESNEFITGMPASTETAYVSSESPIRIS 295
Qy      237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYT 285
      |: |:| :| :| | || | |:|::| |: :: :: | ||| :|
Db      296 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVCVNGGECFTVKDLSNPSRYLCKCPNEFT 355
Qy      286 GDRCQQFAMVNF 297
      ||||| : | :|
Db      356 GDRCQNYVMASF 367

```

RESULT 2

Q9ESA1

```

ID   Q9ESA1          PRELIMINARY;          PRT;          342 AA.
AC   Q9ESA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glial growth factor GGF beta 4 (Fragment).
GN   NRG1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RC   TISSUE=Axotomized lumbar dorsal root ganglion/spinal cord;
RA   Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
RA   Frohnert P.W.;
RT   "Structural and Functional Diversity of Glial Growth Factor Isoforms
RT   Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194997; AAG28451.1; -.
DR   HSSP; Q12784; 1HRE.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   EGF-like domain; Immunoglobulin domain.
FT   NON_TER      1          1
FT   NON_TER      342       342
SQ   SEQUENCE      342 AA;  37836 MW;  8BE36FC836553124 CRC64;

```

```

Query Match          30.2%;  Score 476;  DB 11;  Length 342;
Best Local Similarity 34.3%;  Pred. No. 3.1e-34;
Matches 107;  Conservative 56;  Mismatches 87;  Indels 62;  Gaps 10;

```

Qy 43 VQGLVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92
 | :| || :: :| |||| | :|||::: ::| |
 Db 5 VNWTLPTGGPEPST--DQPGDPAPYLKVHVQVWAVKAGGLKKDSLTVRLDTWGHPAFPS 62

Qy 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKL 147
 | |: : ||||:| | |: :| ||:| |:||||| :|| || |:|
 Db 63 CGRLKEDSRYIFFMEPDANSSGRAPPAFRASFPLET-GRNLKKEVSRVLCKRCALPPRL 121

Qy 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNRKN SRL 204
 |:|||| | | :|| :: :|||:| ||| :||: |: | |
 Db 122 KEMKSQESAAGSKLVLRCESSSEYSSLRFKWFKNGNELNRKNKPNENIKIQKKPGK--SEL 179

Qy 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236
 : || : |:||||: : || : : ||
 Db 180 RINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSESPIRIS 239

Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYT 285
 |: |:| :| :| | || | |:||||| |: :: :: | ||| :|
 Db 240 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVCVNGGECFTVKDLNPSRYLCKCPNEFT 299

Qy 286 GDRCQQFAMVNF 297
 ||||| : | :|
 Db 300 GDRCQNYVMASF 311

RESULT 3

Q9ESA2

ID Q9ESA2 PRELIMINARY; PRT; 323 AA.
 AC Q9ESA2;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Glial growth factor GGF beta 3 (Fragment).
 GN NRG1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
 RA Frohnert P.W.;
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF194996; AAG28450.1; -.
 DR HSSP; Q12784; 1HRE.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.

DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW EGF-like domain; Immunoglobulin domain.
 FT NON_TER 1 1
 SQ SEQUENCE 323 AA; 35358 MW; C7DF153A939A80C8 CRC64;

Query Match 30.1%; Score 473; DB 11; Length 323;
 Best Local Similarity 34.3%; Pred. No. 5.4e-34;
 Matches 107; Conservative 55; Mismatches 88; Indels 62; Gaps 10;

Qy 43 VQGLVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92
 | :| || :: :| |||| |::|||::: ::| |
 Db 5 VNWTLP TGGPEPST--DQPGDPAPYLKVHVQVWAVKAGGLKKDSLTLTVRLDTWGHPAFPS 62
 Qy 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKL 147
 | : : |||||:| | : :| ||:| |:||||| ::|| || :|
 Db 63 CGRLKEDSR YIFFMEPDANSSGRAPPAFRASFP PLET-GRNLKKEVSRVLCKRCALPPRL 121
 Qy 148 KMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNRKNRNSRL 204
 |:|||| | | :|| :: :|||:| |||| :||: | : |
 Db 122 KEMKSQESAAGSKLVLCRCESSSEYSSLRFKWFKNGNELNRKNKPNIKIQKKPGK--SEL 179
 Qy 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236
 : || |:|||:: : || : : ||
 Db 180 RINKASPADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSESPIRIS 239
 Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYT 285
 |: |:| :| :| | || | |:||||| |: :: :: | ||| :|
 Db 240 VSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVCVNGGECFTVKDLNPSRYLCKCPNEFT 299
 Qy 286 GDRCQQFAMVNF 297
 ||||| : | :|
 Db 300 GDRCQNYVMASF 311

RESULT 4

Q9ESA3

ID Q9ESA3 PRELIMINARY; PRT; 317 AA.
 AC Q9ESA3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Glial growth factor GGF beta 2 (Fragment).
 GN NRG1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RC TISSUE=Axotomized lumbar dorsal root ganglion/spinal cord;
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
 RA Frohnert P.W.;
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF194995; AAG28449.1; -.
 DR HSSP; Q12784; 1HRE.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000886; ER_target_S.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW EGF-like domain; Immunoglobulin domain.
 FT NON_TER 1 1
 FT NON_TER 317 317
 SQ SEQUENCE 317 AA; 34785 MW; 4487FA3E9CD876B9 CRC64;

Query Match 29.9%; Score 471; DB 11; Length 317;
 Best Local Similarity 34.0%; Pred. No. 7.9e-34;
 Matches 106; Conservative 57; Mismatches 87; Indels 62; Gaps 10;

Qy 43 VQGLVPAGGSSSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISV-----GS 92
 | :| || :: :| ||| | ::|||::: ::| |
 Db 5 VNWTLP TGGPEPST--DQPGDPAPYLVKVHVQVWAVKAGGLKKDSLLTVRLDTWGHPAFPS 62
 Qy 93 CVPLERNQRYIFFLEPT-----EQPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKL 147
 | |: : ||||:| | |: :| ||:| |:||||| ::|| || |:|
 Db 63 CGRLKEDSR YIFFMEPDANSSGRAPPAFRASFP PLET-GRDLKKEVSRVLCRKRCALPPRL 121
 Qy 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRS---RDIRIKYGNRKNRSRL 204
 |:|||| | | :|| :: :|||:| ||| :||: | : | |
 Db 122 KEMKSQESAAGSKLVLRCE TSSEYSSLRFKWFKNELNRKNKPENIKIQKKPGK--SEL 179
 Qy 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----- 236
 : || : |:| |:| : | | | : | : ||
 Db 180 RINKASLADSGEYMCKVISKLGND SASANITIVESNEFITGMPASTETAYVSSESPIRIS 239
 Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCPVGYT 285
 |: |:| :| :| | || | |:||||| |: :: :| ||| :|
 Db 240 VSTEGANTSSSTSTSTTG TSHLIKCAEKEKTF CVNGGECFTVKDL SNPSRYLCKCPNEFT 299
 Qy 286 GDRCQQFAMVNF 297
 |||| : | :|
 Db 300 GDRCQNYVMASF 311

RESULT 5

Q8NFN3

ID Q8NFN3 PRELIMINARY; PRT; 348 AA.
 AC Q8NFN3;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Neuregulin 1 isoform GGF2 (Fragment).
 GN NRG1.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stefansson H., Sigurdsson E., Steinthorsdottir V., Bjornsdottir S.,
 RA Sigmundsson T., Ghosh S., Brynjolfsson J., Gunnarsdottir S.,
 RA Ivarsson O., Chou T.T., Hjaltason O., Birgisdottir B., Jonsson H.,
 RA Gudnadottir V.G., Gudmundsdottir E., Bjornsson A., Ingvarsson B.,
 RA Ingason A., Sigfusson S., Hardardottir H., Harvey R.P., Brunner D.,
 RA Mutel V., Gonzalo A., Lemke G., Sainz J., Johannesson G.,
 RA Andresson T., Gudbjartsson D., Manolescu A., Frigge M.L., Gurney M.E.,
 RA Kong A., Gulcher J.R., Petursson H., Stefansson K.;
 RT "Neuregulin 1 and susceptibility to Schizophrenia.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF491780; AAM71140.1; -.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00047; ig; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Immunoglobulin domain.
 FT NON_TER 348 348
 SQ SEQUENCE 348 AA; 36997 MW; 15568C6260C5635C CRC64;

Query Match 25.7%; Score 404.5; DB 4; Length 348;
 Best Local Similarity 34.4%; Pred. No. 8.1e-28;
 Matches 100; Conservative 49; Mismatches 69; Indels 73; Gaps 11;

Qy 15 GVSLACYS--PSLKSVQDQAYKAPVVVEGKV-----QGLV-----PAGGSSS--NSTRE 59
 | | : ||| ||: |||: | :| ||: ||| | | : | | : ||
 Db 58 GASV-CYSSPPSVGSGVQELAQRAAVVIEGKVHPQRRQQGALDRKAAAAAGEAGAWGGDRE 116
 Qy 60 PPASGRVA-----LVKVLDKWPLRSGGLQ 83
 |||:| | ||| | :|||:|
 Db 117 PPAAGPRALGPPAEELPLAANGTVPSWPTAPVPSAGEPGEEAPYLKVHVQVWAVKAGGLK 176
 Qy 84 REQVISV-----GSCVPLERNQRYIFFLEP-----TEQPLVFKTAFAPLDTNGKN 128
 :: ::| || | : : ||||: || : | | : : | ||: | :|
 Db 177 KDSLLTVRLGTWGHAPAFPSGRLKEDSRYIFFMEPDANSTSRAPAAFRASFPPLET-GRN 235
 Qy 129 LKKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRS 188
 ||||| :||| || | :||: ||| | | | :|| : : ||||: | |||
 Db 236 LKKEVSRVLCRKRCALPPRLKEMKSQESAAGSKLVLRCEISSEYSSLRFKWFKNGNELNRK 295
 Qy 189 ---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYV 236
 ::||: | : | | : | : ||||: | : | | : :
 Db 296 NKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSANITI 344

RESULT 6

Q07112

ID Q07112 PRELIMINARY; PRT; 241 AA.

AC Q07112;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Glial growth factor.
 GN GGFBBP5.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Posterior pituitary;
 RX MEDLINE=93205115; PubMed=8096067;
 RA Marchionni M.A., Goodearl A.D.G., Chen M., Bermingham-McDonogh O.,
 RA Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J.,
 RA Kobayashi K., Wroblewski D., Lynch C., Baldassarre M., Hiles I.,
 RA Davis J.B., Hsuan J., Totty N.F., Otsu M., McBurney R.N.,
 RA Waterfield M.D., Stroobant P., Gwynne D.;
 RT "Glial growth factors are alternatively spliced erbB2 ligands
 RT expressed in the nervous system.";
 RL Nature 362:312-318(1993).
 DR EMBL; L12259; AAA30540.1; -.
 DR PIR; S32359; S32359.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50835; IG_LIKE; 1.
 SQ SEQUENCE 241 AA; 25955 MW; BF571297E8DA9796 CRC64;

Query Match 18.6%; Score 292; DB 6; Length 241;
 Best Local Similarity 32.8%; Pred. No. 5.9e-18;
 Matches 65; Conservative 37; Mismatches 52; Indels 44; Gaps 6;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKEL---NRSRDIRIKYNG 198
 | | :|:|:| | | | | :|:| : : :|:|:| | | | : : :|:|:| |
 Db 34 ALPPRLKEMKSQESVAGSKLVLRCESSSEYSSLKFKWFKNGSELSRKNKPQNIKIQRPG 93

 Qy 199 RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV---- 236
 : | | : : | : | :|:|:|:| : | | : : | |
 Db 94 K--SELRISKASLADSGEYMCKVISKLGNDASANITIVESNEITTGMPASTETAYVSSE 151

 Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCCYYIEGINQLS---CK 279
 | : | : | : | : | | | | : :|:|:| | | : : : : | | |
 Db 152 SPIRISVSTEGTNTSSSTSTSTAGTSHLVKCAEKEKTFVNGGECFMVKDLNPSRYLCK 211

 Qy 280 CPVGYTGDRCCQFAMVNF 297
 | | :|:|:| | : | : |
 Db 212 CPNEFTGDRCCQNYVMASF 229

RESULT 7

O35947

ID O35947 PRELIMINARY; PRT; 461 AA.
AC O35947;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pro-neuregulin-1, isoform alpha 2B precursor.
GN NRG1 OR NDF.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA2B), AND SEQUENCE OF 64-81.
RC TISSUE=EMBRYO;
RX MEDLINE=98196996; PubMed=9537646;
RA Velasco J.A., Feijoo E., Avila M.A., Notario V.;
RT "Secretion of neu differentiation factor-like polypeptides by cph-
RT transformed fibroblasts: cloning and characterization of Syrian
RT hamster neuregulin cDNAs.";
RL Mol. Carcinog. 21:156-163(1998).
CC -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE
CC RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,
CC RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND
CC ACTIVATION OF THE ERBB RECEPTORS. MAY PLAY AN IMPORTANT ROLE IN
CC PROVIDING GROWTH ADVANTAGE IN NEOPLASTIC CELLS.
CC -!- SUBUNIT: THE CYTOPLASMIC DOMAIN INTERACTS WITH THE LIM DOMAIN
CC REGION OF LIMK1 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: EXISTS AS TYPE I MEMBRANE PROTEIN AND AS A
CC PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE MEMBRANE-
CC BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHER LEVEL AFTER NEOPLASMIC
CC TRANSFORMATION OF CELLS.
CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION
CC OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE
CC PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN
CC DIMERIZATION (BY SIMILARITY).
CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE
CC DOMAIN.
CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE
CC EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR
CC FORM (BY SIMILARITY).
CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES PROTEOLYTIC CLEAVAGE (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.
DR EMBL; U96612; AAB71812.1; -.
DR HSSP; Q12784; 1HRE.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008083; F:growth factor activity; IEA.
DR GO; GO:0009790; P:embryonic development; IEA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.

DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002154; Neuregulin.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF02158; Neuregulin; 1.
 DR PRINTS; PR01089; NEUREGULIN.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 KW Transmembrane; Alternative splicing.
 FT PROPEP 1 13 BY SIMILARITY.
 FT CHAIN 14 461 PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.
 FT CHAIN 14 241 NEUREGULIN-1.
 FT DOMAIN 14 242 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 243 265 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
 FT DOMAIN 266 461 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 50 119 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 165 177 SER/THR-RICH.
 FT DOMAIN 178 222 EGF-LIKE.
 FT DISULFID 57 112 BY SIMILARITY.
 FT DISULFID 182 196 BY SIMILARITY.
 FT DISULFID 190 210 BY SIMILARITY.
 FT DISULFID 212 221 BY SIMILARITY.
 FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 461 AA; 50890 MW; 935C9560F7148336 CRC64;

Query Match 17.6%; Score 277; DB 11; Length 461;

Best Local Similarity 33.2%; Pred. No. 3.1e-16;

Matches 63; Conservative 31; Mismatches 56; Indels 40; Gaps 5;

Qy 142 ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELN-RSRDIRIKYGNGRK 200
 | |:|:| | | | |:|:|: |: |:|:|:| | | |:|:| |
 Db 34 ALPPRLKEMKIQESAAGSKLVLCETSSSEYPELRFKWFKNSELNKRTPQNIKLQKKPG 93
 Qy 201 NSRLQFNKVKVEDAGEYVCEAENILGKDTVGRG-----RLYV----- 236
 | |:|:| |:|:|:|: |:| |:| | | |
 Db 94 KSELRINKASLADSGEYMCKVISKLGNDASANITIVDSNEFITGMPASTERAYVSSSESP 153
 Qy 237 -----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGINQLS---CKCP 281
 |: |:|:| |:| | | | |:|:|:| |: |:|:| | | |
 Db 154 IRISVSTEGANTSSSTSTSTTGTSHLVKCAEKEKTFVNGGECFMVKDLSNPSRYLCKCQ 213
 Qy 282 VGYTGDRCCQ 291
 |:|:| |:|:|
 Db 214 PGFTGARCTE 223

RESULT 8

Q810X1

ID Q810X1 PRELIMINARY; PRT; 54 AA.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RA Carroll S.L., Stonecypher M.S., Anderson K.D., Pearson R.J. Jr.,
 RA Frohnert P.W.;
 RT "Structural and Functional Diversity of Glial Growth Factor Isoforms
 RT Expressed in Regenerating Peripheral Nerve and Associated Neurons.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF194994; AAG28434.1; -.
 FT NON_TER 244 244
 SQ SEQUENCE 244 AA; 25866 MW; 019CBC2DFFF8F625 CRC64;

Query Match 14.0%; Score 221; DB 11; Length 244;
 Best Local Similarity 29.5%; Pred. No. 1.4e-11;
 Matches 62; Conservative 28; Mismatches 44; Indels 76; Gaps 8;

Qy 5 PAPGFSMLLFGVSL-----ACYS--PSLKSVDQAYKAPVVVEGKV--- 43
 | | :| | : || | | :| | :| | :| | :| |
 Db 38 PPPLLLLLLLGTAALAPGAAAERAAPAGASVCYSSPPSVGSGVQELARRAAVVEGKVHPP 97
 Qy 44 ---QG-----LVPAGGSSSNSTREPPASGRV 66
 || :| | : :|
 Db 98 RRQQGALDRKAAGEAGAGARDQPVQDSPPSQDPLPAVNWTLPTGGPEPST--DQPGDPAP 155
 Qy 67 ALVKVLDKWPLRSGGLQREQVISV-----GSCVPLERNQRYIFFLEPT-----EQ 111
 ||| | : :| | : : :| | | : :| | | :| |
 Db 156 YLVKVHVQVAVKAGGLKKDSLTLVRLDTWGHPAFPSCGRLKEDSRYIFFMEPDANSSGRA 215
 Qy 112 PLVFKTAFAPLDTNGKNLKKEVGKILCTDC 141
 | | : :| | :| | :| | | :| | |
 Db 216 PPAFRASFPPLET-GRNLKKEVSRVLCKRC 244

RESULT 11

Q810X2

ID Q810X2 PRELIMINARY; PRT; 79 AA.
 AC Q810X2;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Neuregulin 2-alpha (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Olfactory bulb;
 RA Mautino B., Dalla Costa L., Dati C.;
 RT "Bioactive recombinant NRG1, NRG2 and NRG3 expressed in E. coli.";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY227025; AAO72522.1; -.
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00008; EGF; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 FT NON_TER 1 1

SQ SEQUENCE 79 AA; 8727 MW; DA4501900C610780 CRC64;

Query Match 12.7%; Score 200; DB 11; Length 79;
Best Local Similarity 89.5%; Pred. No. 2.3e-10;
Matches 34; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 252 KCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRC 289

||||| : | |

Db 1 KCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 38

RESULT 12

Q8NFN2

ID Q8NFN2 PRELIMINARY; PRT; 167 AA.

AC Q8NFN2;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Neuregulin 1 isoform GGF (Fragment).

GN NRGL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Stefansson H., Sigurdsson E., Steinthorsdottir V., Bjornsdottir S.,

RA Sigmundsson T., Ghosh S., Brynjolfsson J., Gunnarsdottir S.,

RA Ivarsson O., Chou T.T., Hjaltason O., Birgisdottir B., Jonsson H.,

RA Gudnadottir V.G., Gudmundsdottir E., Bjornsson A., Ingvarsson B.,

RA Ingason A., Sigfusson S., Hardardottir H., Harvey R.P., Brunner D.,

RA Mutel V., Gonzalo A., Lemke G., Sainz J., Johannesson G.,

RA Andresson T., Gudbjartsson D., Manolescu A., Frigge M.L., Gurney M.E.,

RA Kong A., Gulcher J.R., Petursson H., Stefansson K.;

RT "Neuregulin 1 and susceptibility to Schizophrenia.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF491780; AAM71139.1; -.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003598; Ig_c2.

DR Pfam; PF00047; ig; 1.

DR SMART; SM00409; IG; 1.

DR SMART; SM00408; IGc2; 1.

DR PROSITE; PS50835; IG_LIKE; 1.

KW Immunoglobulin domain.

FT NON TER 167 167

SQ SEQUENCE 167 AA; 17983 MW; 9C2FB3A579325FF4 CRC64;

Query Match 11.5%; Score 180.5; DB 4; Length 167;
Best Local Similarity 30.9%; Pred. No. 3.5e-08;
Matches 51; Conservative 30; Mismatches 63; Indels 21; Gaps 5;

Qy 126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178

|| ||| | : | :||:|||| | | |:| : :|

Db 11 GKGGKKKRGSGKKPESAAGSQSPALPPRLKEMKQESAAGSKLVLCETSSEYSSLRFKW 70

Qy 179 FKDGKELNRS---RDIRIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLY 235

```

      ||:| |||| :||:| :| | :| | :|:||||:| :|| |: :
Db      71 FKNGNELNRKNKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSASANIT 128

Qy      236 V---NSVSTTLSSWSGHARKCNETAKSYCVNGGVVCYYIEGINQLS 277
      : | : | : : : | :|: | : | | | |
Db      129 IVESNEIITGMPASTEGAYVSSESPIRISVS-----TEGANTSS 167

```

RESULT 13

Q86GD6

```

ID   Q86GD6          PRELIMINARY;      PRT;   8625 AA.
AC   Q86GD6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Projectin.
GN   PROJ.
OS   Procambarus clarkii (Red swamp crayfish).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Astacoidea; Cambaridae; Procambarus.
OX   NCBI_TaxID=6728;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RA   Oshino T., Shimamura J., Fukuzawa A., Maruyama K., Kimura S.;
RT   "The entire cDNA sequences of projectin isoforms of crayfish claw
RT   closer and flexor muscles and their localization.";
RL   J. Muscle Res. Cell. Motil. 0:0-0(2003).
DR   EMBL; AB055927; BAC66140.1; -.
DR   GO; GO:0005743; C:mitochondrial inner membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0005488; F:binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0003743; F:translation initiation factor activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0006413; P:translational initiation; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001950; TIF_SUI1.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 39.
DR   Pfam; PF00047; ig; 13.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Prot_kinase; 1.

```

DR SMART; SM00060; FN3; 39.
 DR SMART; SM00409; IG; 36.
 DR SMART; SM00408; IGc2; 24.
 DR SMART; SM00406; IGv; 3.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS50835; IG_LIKE; 24.
 DR PROSITE; PS00215; MITOCH_CARRIER; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS01118; SUI1_1; 1.
 SQ SEQUENCE 8625 AA; 962637 MW; 56B8E4C4FE0AFC90 CRC64;

Query Match 10.5%; Score 166; DB 5; Length 8625;
 Best Local Similarity 24.2%; Pred. No. 0.00014;
 Matches 87; Conservative 39; Mismatches 130; Indels 104; Gaps 16;

Qy 2 RRDP-----APGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPA 49
 || | || | : | : | | : || ||
 Db 8115 RRQPLTYKLWQEEGEGAPSFTFLLRPRVIQCH-----QTCKLLCCLAGKP---VPT 8162
 Qy 50 ----GGSSSNST---REPPASGRVALVKVLDKWPLRSGGLQREQVISVG----SCVPLER 98
 || | : | | | : : : | | : | : || :
 Db 8163 VKWYKGSQELSKFDYSQSHADG-VVTIEIVNCKPADSGKYRCVATNSLGTDETSCVVIVE 8221
 Qy 99 NQRYIFF----LEPTEQPLV-----FKTAFAPLDTNGK---NLKKEVGKILCTDCATRP 145
 : : || | | | : : | : : | : | | |
 Db 8222 DRRYIETTIKDLPPPPTPAIRVDDTSSSSYFTSTHKDGRSSTSTKVEAASSSSTSSAAAS 8281
 Qy 146 KLK-----KMKSQTGQV-----GEKQS 162
 | : | | ||
 Db 8282 GAKRTLKPYGKRQDSTGSTRSRSATKELELPPDDSLMGPPGFSGELPKTLAIKDGEALC 8341
 Qy 163 LKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKNRSLQFNKVVEDAGEYVCEAE 222
 ||| | : : | | || | : | : | | : | | : |
 Db 8342 LKC-TVKGDPEPQVSWFKDGEPLSSSDIIDLKYRQGL--ASLTINEVFPEDGLYVCKAT 8398
 Qy 223 NILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAK---SYCVNGGVCYYIEGINQLSCK 279
 : || : : | : : || | : | | : |||
 Db 8399 SSLGSAETKCKLSISPMEQQINGKSGRGDKLPRITQHLLSQEVPDGTAH-----TLCK 8452

RESULT 14

Q8I0L3

ID Q8I0L3 PRELIMINARY; PRT; 5175 AA.
 AC Q8I0L3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE C. elegans him-4 protein (corresponding sequence F15G9.4a).
 GN F15G9.4 OR HIM-4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Sulston J.E.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kershaw J.K.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; Z47068; CAA87335.1; -.
 DR EMBL; Z47070; CAA87335.1; JOINED.
 DR EMBL; Z47070; CAA87344.1; -.
 DR EMBL; Z47068; CAA87344.1; JOINED.
 DR PIR; T20992; T20992.
 DR WormPep; F15G9.4a; CE18595.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000515; BPD_transp.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR003596; Ig_v.
 DR Pfam; PF00047; ig; 47.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00409; IG; 45.
 DR SMART; SM00408; IGc2; 47.
 DR SMART; SM00406; IGv; 12.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR PROSITE; PS50835; IG_LIKE; 47.
 SQ SEQUENCE 5175 AA; 568471 MW; 4B2561803BBC62A4 CRC64;

Query Match 10.5%; Score 164.5; DB 5; Length 5175;
 Best Local Similarity 26.5%; Pred. No. 9.3e-05;
 Matches 62; Conservative 26; Mismatches 79; Indels 67; Gaps 10;

Qy 49 AGGSSSNSTR---EPPASGRVALVKVLDK-----WPLRS 79
 ||| |: | |||: ||| : | :
 Db 595 AGGMSTRKMRLDIMEPPS-----VKVTPQDVYFNMREGVNLSCEAMGDPKPEVHWYFKG 648
 Qy 80 GGLQREQVISVGSCVPLERNQRYIFFL-----EPTEQPLVFKTAFAPLDTNGKNLKKEV 133
 | : || :: :::: || : | || |
 Db 649 RHLLNDYKYQVG-----QDSKFLYIRDATHHDEGTYECRAMSQAGQARDTTDLML---- 698

QY 134 GKILCTDCATRPKLK--KMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDI 191
 || ||:: : | | : |:: | :: | | | |||:|:| : |
 Db 699 -----ATPPKVEIIQNKMVGR-GDRVSECKTIRGKPHPKIRWFKNGKDLIKPDDY 749
 QY 192 RIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSS 245
 || | :| | ||| | | ||: || | | | : |
 Db 750 -IKINEG---QLHIMGAKDEDAGAYSCVGENMAGKDVQVANLSVGRVPTIIES 798

RESULT 15

O76518

ID O76518 PRELIMINARY; PRT; 5198 AA.
 AC O76518; Q10036;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hemicentin precursor (C. elegans him-4 protein) (corresponding
 DE sequence F15G9.4b).
 GN F15G9.4 OR HIM-4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX PubMed=11222143;
 RA Vogel B.E., Hedgecock E.M.;
 RT "Hemicentin, a conserved extracellular member of the immunoglobulin
 RT superfamily, organizes epithelial and other cell attachments into
 RT oriented line-shaped junctions.";
 RL Development 128:883-894(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Sulston J.E.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Kershaw J.K.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF074901; AAC26792.1; -.
 DR EMBL; Z47068; CAA87336.1; -.
 DR EMBL; Z47070; CAA87336.1; JOINED.
 DR EMBL; Z47070; CAA87345.1; -.
 DR EMBL; Z47068; CAA87345.1; JOINED.
 DR PIR; T43290; T43290.
 DR HSSP; P00736; 1APQ.
 DR WormPep; F15G9.4b; CE18596.
 DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR000515; BPD_transp.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00047; ig; 47.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00408; IGc2; 44.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR PROSITE; PS50835; IG_LIKE; 47.
 KW EGF-like domain; Immunoglobulin domain; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 5198 HEMICENTIN.
 SQ SEQUENCE 5198 AA; 570809 MW; DA8511FF2B58D37B CRC64;

Query Match 10.5%; Score 164.5; DB 5; Length 5198;
 Best Local Similarity 26.5%; Pred. No. 9.4e-05;
 Matches 62; Conservative 26; Mismatches 79; Indels 67; Gaps 10;

Qy 49 AGGSSSNSTR---EPPASGRVALVKVLDK-----WPLRS 79
 ||| |: | ||: ||| : | :
 Db 595 AGGMSTRKMRLDIMEPPS-----VKVTPQDVYFNMREGVNLSCAMGDPKPEVHWYFKG 648

 Qy 80 GGLQREQVISVGSVCVPLERNQRYIFFL-----EPTEQPLVFKTAFAPLDTNGKNLKKEV 133
 | : || ::::: | | : | || |
 Db 649 RHLLNDYKYQVG-----QDSKFLYIRDATHHDEGTYECRAMSQAGQARDTTDLML----- 698

 Qy 134 GKILCTDCATRPKLK--KMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDI 191
 || ||:: : | | : |:: | :|: | | | |||:|:| : |
 Db 699 -----ATPPKVEIIQNKMVGR-GDRVSFECKTIRGKPHPKIRWFKNGKDLIKPDDY 749

 Qy 192 RIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRRLYVNSVSTTLSS 245
 || | :| | ||| | | ||: || | | | : |
 Db 750 -IKINEG----QLHIMGAKDEDAGAYSCVGENMAGKDVQVANLSVGRVPTIIES 798

Search completed: August 17, 2004, 14:12:41
 Job time : 34.6911 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 13:56:40 ; Search time 8.5414 Seconds
(without alignments)
1816.670 Million cell updates/sec

Title: US-09-864-675-4
Perfect score: 1574
Sequence: 1 MRRDPAPGFSMLLFGVSLAC.....KCPVGYTGDRCQQFAMVNFS 298

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1544	98.1	756	1	NRG2_MOUSE	P56974 mus musculu
2	1505	95.6	850	1	NRG2_HUMAN	O14511 homo sapien
3	1470	93.4	868	1	NRG2_RAT	O35569 rattus norv
4	305.5	19.4	602	1	NRG1_CHICK	Q05199 gallus gall
5	299.5	19.0	677	1	NRG1_XENLA	O93383 xenopus lae
6	299	19.0	662	1	NRG1_RAT	P43322 r pro-neure
7	280.5	17.8	639	1	NRG1_HUMAN	Q02297 h pro-neure
8	208.5	13.2	623	1	VEIN_DROME	Q94918 drosophila
9	155	9.8	298	1	JAM2_HUMAN	P57087 homo sapien
10	152	9.7	296	1	SMDF_HUMAN	Q15491 homo sapien
11	150	9.5	1217	1	EGF_MOUSE	P01132 mus musculu
12	145.5	9.2	2012	1	DSCA_HUMAN	O60469 homo sapien
13	145	9.2	338	1	LAMP_RAT	Q62813 rattus norv
14	144	9.1	338	1	LAMP_HUMAN	Q13449 homo sapien
15	141.5	9.0	353	1	CEPU_CHICK	Q90773 gallus gall
16	135.5	8.6	1356	1	VGR2_HUMAN	P35968 homo sapien
17	135	8.6	4391	1	PGBM_HUMAN	P98160 homo sapien

18	133.5	8.5	1367	1	VGR2_MOUSE	P35918	mus musculu
19	133.5	8.5	6632	1	UN89_CAEEL	O01761	caenorhabdi
20	132.5	8.4	1133	1	EGF_RAT	P07522	rattus norv
21	131.5	8.4	338	1	LAMP_CHICK	Q98919	gallus gall
22	131.5	8.4	1343	1	VGR2_RAT	O08775	rattus norv
23	131	8.3	345	1	OPCM_RAT	P32736	rattus norv
24	130.5	8.3	824	1	MLT1_HUMAN	Q9udy8	homo sapien
25	129.5	8.2	837	1	NCM2_MOUSE	O35136	mus musculu
26	129.5	8.2	1018	1	CONT_HUMAN	Q12860	homo sapien
27	129.5	8.2	1064	1	FBP1_STRPU	P10079	strongyloce
28	129	8.2	345	1	OPCM_HUMAN	Q14982	homo sapien
29	128.5	8.2	1010	1	CONT_CHICK	P14781	gallus gall
30	128.5	8.2	1020	1	CONT_MOUSE	P12960	mus musculu
31	128.5	8.2	1021	1	CONT_RAT	Q63198	rattus norv
32	128.5	8.2	1091	1	NCA1_CHICK	P13590	gallus gall
33	128	8.1	1040	1	AXO1_HUMAN	Q02246	homo sapien
34	127.5	8.1	761	1	NCA2_HUMAN	P13592	homo sapien
35	127.5	8.1	848	1	NCA1_HUMAN	P13591	homo sapien
36	127	8.1	298	1	JAM1_BOVIN	Q9xt56	bos taurus
37	127	8.1	344	1	NTRI_RAT	Q62718	rattus norv
38	126.5	8.0	1040	1	AXO1_RAT	P22063	rattus norv
39	125.5	8.0	853	1	NCA1_BOVIN	P31836	bos taurus
40	125.5	8.0	937	1	ROR1_HUMAN	Q01973	homo sapien
41	125	7.9	937	1	ROR1_MOUSE	Q9z139	mus musculu
42	124.5	7.9	1036	1	AXO1_CHICK	P28685	gallus gall
43	124	7.9	344	1	NTRI_MOUSE	Q99pj0	mus musculu
44	123	7.8	345	1	OPCM_BOVIN	P11834	bos taurus
45	122	7.8	53	1	EGF_PIG	Q00968	sus scrofa

ALIGNMENTS

RESULT 1

NRG2_MOUSE

ID NRG2_MOUSE STANDARD; PRT; 756 AA.
AC P56974;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)
DE (Divergent of neuregulin 1) (DON-1)].
GN NRG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS NRG2-5; NRG2-10 AND NRG2-16A).
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=97311398; PubMed=9168115;
RA Carraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N.,
RA Gassmann M., Lai C.;
RT "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine
RT kinases.";
RL Nature 387:512-516(1997).
RN [2]

RP SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).
 RC TISSUE=Choroid plexus;
 RX MEDLINE=97342638; PubMed=9199335;
 RA Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,
 RA Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,
 RA Gearing D.P.;
 RT "Characterization of a neuregulin-related gene, Don-1, that is highly
 RT expressed in restricted regions of the cerebellum and hippocampus.";
 RL Mol. Cell. Biol. 17:4007-4014(1997).
 CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase
 CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,
 CC resulting in ligand-stimulated tyrosine phosphorylation and
 CC activation of the ERBB receptors. May also promote the
 CC heterodimerization with the EGF receptor.
 CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS
 CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
 CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=NRG2-16A;
 CC IsoId=P56974-1; Sequence=Displayed;
 CC Name=DON-1M;
 CC IsoId=P56974-2; Sequence=VSP_003464;
 CC Name=DON-1S; Synonyms=NRG2-5;
 CC IsoId=P56974-3; Sequence=VSP_003462, VSP_003463;
 CC Name=NRG2-10;
 CC IsoId=P56974-4; Sequence=VSP_003460, VSP_003461;
 CC -!- TISSUE SPECIFICITY: Highest expression in the brain, with lower
 CC levels in the lung. In the cerebellum, found in granule and
 CC Purkinje cells.
 CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
 CC of trafficking and proteolytic processing. Regulation of the
 CC proteolytic processing involves initial intracellular domain
 CC dimerization (By similarity).
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
 CC domain (By similarity).
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the
 CC external face leads to the release of the soluble growth factor
 CC form (By similarity).
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
 CC similarity).
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
 CC -!- SIMILARITY: Belongs to the neuregulin family.
 DR HSSP; Q12784; 1HRE.
 DR MGD; MGI:1098246; Nrg2.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002154; Neuregulin.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF02158; Neuregulin; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.

DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50026; EGF_3; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 KW Transmembrane; Multigene family; Alternative splicing.
 FT PROPEP 1 19 BY SIMILARITY.
 FT CHAIN 20 756 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
 FT CHAIN 20 314 NEUREGULIN-2.
 FT DOMAIN 20 315 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 316 336 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
 FT DOMAIN 337 756 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 145 240 IG-LIKE C2-TYPE.
 FT DOMAIN 238 248 SER/THR-RICH.
 FT DOMAIN 249 290 EGF-LIKE.
 FT DOMAIN 627 633 POLY-PRO.
 FT DISULFID 165 219 BY SIMILARITY.
 FT DISULFID 253 267 BY SIMILARITY.
 FT DISULFID 261 278 BY SIMILARITY.
 FT DISULFID 280 289 BY SIMILARITY.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 280 280 C -> G (in isoform NRG2-10).
 FT /FTId=VSP_003460.
 FT VARSPLIC 281 756 Missing (in isoform NRG2-10).
 FT /FTId=VSP_003461.
 FT VARSPLIC 282 330 VGYTGDRCCQFAMVNFSKHLGFELKEAEELYQKRVLITITGI
 FT CVALLVVG -> NGFFGQRCLEKLPLRLYMPDPKQSVLWDT
 FT PGTGVSSSQWSTSPSTLDLN (in isoform DON-1S).
 FT /FTId=VSP_003462.
 FT VARSPLIC 331 756 Missing (in isoform DON-1S).
 FT /FTId=VSP_003463.
 FT VARSPLIC 282 307 VGYTGDRCCQFAMVNFSKHLGFELKE -> NGFFGQRCLEK
 FT LPLRLYMPDPKQK (in isoform DON-1M).
 FT /FTId=VSP_003464.
 SQ SEQUENCE 756 AA; 82213 MW; 51D85DC918BE678E CRC64;

Query Match 98.1%; Score 1544; DB 1; Length 756;

Best Local Similarity 97.3%; Pred. No. 5e-116;

Matches 290; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
 |||||
 Db 1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 60
 Qy 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFA 120
 |||||
 Db 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFA 120
 Qy 121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
 |:| ||||:|||||
 Db 121 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQGTGEVGEKQSLKCEAAAGNPQPSYRWFK 180
 Qy 181 DGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGRLYVNSVS 240
 |||||:|||||

Db 181 DGKELNRSRDIRIKYGNRKNRLQFNKVRVEDAGEYVCEAENILGKDTVGRRLHVNVS 240

Qy 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCQQFAMVNFS 298
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRCQQFAMVNFS 298

RESULT 2

NRG2_HUMAN

ID NRG2_HUMAN STANDARD; PRT; 850 AA.

AC O14511;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)

DE (Neural-and thymus-derived activator for ERBB kinases) (NTAK)

DE (Divergent of neuregulin 1) (DON-1)].

GN NRG2 OR NTAK.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Neuroblastoma;

RX MEDLINE=98006324; PubMed=9348101;

RA Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,

RA Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,

RA Ishiguro H.;

RT "A novel brain-derived member of the epidermal growth factor family

RT that interacts with ErbB3 and ErbB4.";

RL J. Biochem. 122:675-680(1997).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS DON-1B AND DON-1R).

RC TISSUE=Fetal brain;

RX MEDLINE=97342638; PubMed=9199335;

RA Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,

RA Woolf E.A., Comrack C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,

RA Gearing D.P.;

RT "Characterization of a neuregulin-related gene, Don-1, that is highly

RT expressed in restricted regions of the cerebellum and hippocampus.";

RL Mol. Cell. Biol. 17:4007-4014(1997).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).

RC TISSUE=Fetal brain, and Lung;

RX MEDLINE=99295836; PubMed=10369162;

RA Ring H.Z., Chang H., Guilbot A., Brice A., LeGuern E., Francke U.;

RT "The human neuregulin 2 (NRG2) gene: cloning, mapping and evaluation

RT as a candidate for the autosomal recessive form of Charcot-Marie-Tooth

RT disease linked to 5q.";

RL Hum. Genet. 104:326-332(1999).

CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase

CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,

CC resulting in ligand-stimulated tyrosine phosphorylation and

CC activation of the ERBB receptors. May also promote the

CC heterodimerization with the EGF receptor.

CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS

```

CC      A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE
CC      MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).
CC      -!- ALTERNATIVE PRODUCTS:
CC          Event=Alternative splicing; Named isoforms=8;
CC          Name=1;
CC              IsoId=O14511-1; Sequence=Displayed;
CC          Name=2;
CC              IsoId=O14511-2; Sequence=VSP_003453;
CC          Name=3;
CC              IsoId=O14511-3; Sequence=VSP_003455;
CC          Name=4;
CC              IsoId=O14511-4; Sequence=VSP_003454;
CC          Name=5;
CC              IsoId=O14511-5; Sequence=VSP_003458, VSP_003459;
CC          Name=6;
CC              IsoId=O14511-6; Sequence=VSP_003456, VSP_003457;
CC          Name=DON-1B;
CC              IsoId=O14511-7; Sequence=VSP_003452, VSP_003455;
CC          Name=DON-1R;
CC              IsoId=O14511-8; Sequence=VSP_003451;
CC      -!- TISSUE SPECIFICITY: Restricted to the cerebellum in the adult.
CC      -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
CC                  of trafficking and proteolytic processing. Regulation of the
CC                  proteolytic processing involves initial intracellular domain
CC                  dimerization (By similarity).
CC      -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
CC                  domain (By similarity).
CC      -!- PTM: Proteolytic cleavage close to the plasma membrane on the
CC                  external face leads to the release of the soluble growth factor
CC                  form (By similarity).
CC      -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC                  similarity).
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AB005060; BAA23417.1; -.
DR      EMBL; AF119162; AAF28848.1; -.
DR      EMBL; AF119151; AAF28848.1; JOINED.
DR      EMBL; AF119152; AAF28848.1; JOINED.
DR      EMBL; AF119153; AAF28848.1; JOINED.
DR      EMBL; AF119154; AAF28848.1; JOINED.
DR      EMBL; AF119155; AAF28848.1; JOINED.
DR      EMBL; AF119158; AAF28848.1; JOINED.
DR      EMBL; AF119159; AAF28848.1; JOINED.
DR      EMBL; AF119160; AAF28848.1; JOINED.
DR      EMBL; AF119161; AAF28848.1; JOINED.
DR      EMBL; AF119162; AAF28849.1; -.
DR      EMBL; AF119151; AAF28849.1; JOINED.

```


DR EMBL; AF119152; AAF28849.1; JOINED.
DR EMBL; AF119153; AAF28849.1; JOINED.
DR EMBL; AF119154; AAF28849.1; JOINED.
DR EMBL; AF119156; AAF28849.1; JOINED.
DR EMBL; AF119158; AAF28849.1; JOINED.
DR EMBL; AF119159; AAF28849.1; JOINED.
DR EMBL; AF119160; AAF28849.1; JOINED.
DR EMBL; AF119161; AAF28849.1; JOINED.
DR EMBL; AF119162; AAF28850.1; -.
DR EMBL; AF119151; AAF28850.1; JOINED.
DR EMBL; AF119152; AAF28850.1; JOINED.
DR EMBL; AF119153; AAF28850.1; JOINED.
DR EMBL; AF119154; AAF28850.1; JOINED.
DR EMBL; AF119155; AAF28850.1; JOINED.
DR EMBL; AF119157; AAF28850.1; JOINED.
DR EMBL; AF119158; AAF28850.1; JOINED.
DR EMBL; AF119159; AAF28850.1; JOINED.
DR EMBL; AF119160; AAF28850.1; JOINED.
DR EMBL; AF119161; AAF28850.1; JOINED.
DR EMBL; AF119162; AAF28851.1; -.
DR EMBL; AF119151; AAF28851.1; JOINED.
DR EMBL; AF119152; AAF28851.1; JOINED.
DR EMBL; AF119153; AAF28851.1; JOINED.
DR EMBL; AF119154; AAF28851.1; JOINED.
DR EMBL; AF119156; AAF28851.1; JOINED.
DR EMBL; AF119157; AAF28851.1; JOINED.
DR EMBL; AF119158; AAF28851.1; JOINED.
DR EMBL; AF119159; AAF28851.1; JOINED.
DR EMBL; AF119160; AAF28851.1; JOINED.
DR EMBL; AF119161; AAF28851.1; JOINED.
DR EMBL; AF119158; AAF28852.1; -.
DR EMBL; AF119151; AAF28852.1; JOINED.
DR EMBL; AF119152; AAF28852.1; JOINED.
DR EMBL; AF119153; AAF28852.1; JOINED.
DR EMBL; AF119154; AAF28852.1; JOINED.
DR EMBL; AF119155; AAF28852.1; JOINED.
DR EMBL; AF119156; AAF28852.1; JOINED.
DR EMBL; AF119157; AAF28853.1; -.
DR EMBL; AF119151; AAF28853.1; JOINED.
DR EMBL; AF119152; AAF28853.1; JOINED.
DR EMBL; AF119153; AAF28853.1; JOINED.
DR EMBL; AF119154; AAF28853.1; JOINED.
DR EMBL; AF119155; AAF28853.1; JOINED.
DR EMBL; AF119156; AAF28853.1; JOINED.
DR PIR; JC5700; JC5700.
DR HSSP; Q12784; 1HRE.
DR Genew; HGNC:7998; NRG2.
DR MIM; 603818; -.
DR GO; GO:0005102; F:receptor binding; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR002154; Neuregulin.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00047; ig; 1.

DR Pfam; PF02158; Neuregulin; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50026; EGF_3; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 KW Transmembrane; Multigene family; Alternative splicing.
 FT PROPEP 1 111 BY SIMILARITY.
 FT CHAIN 112 850 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
 FT CHAIN 112 404 NEUREGULIN-2.
 FT DOMAIN 112 405 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 406 426 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
 FT DOMAIN 427 850 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 237 332 IG-LIKE C2-TYPE.
 FT DOMAIN 330 340 SER/THR-RICH.
 FT DOMAIN 341 382 EGF-LIKE.
 FT DOMAIN 10 13 POLY-PRO.
 FT DOMAIN 20 30 POLY-SER.
 FT DOMAIN 33 47 POLY-SER.
 FT DOMAIN 87 90 POLY-ALA.
 FT DOMAIN 721 727 POLY-PRO.
 FT DISULFID 257 311 BY SIMILARITY.
 FT DISULFID 345 359 BY SIMILARITY.
 FT DISULFID 353 370 BY SIMILARITY.
 FT DISULFID 372 381 BY SIMILARITY.
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 233 MRQVCCSALPPPPLEKGRCSSYSDSSSSSSSSSSSSSSSSSSSSSS
 FT ESGSSSRSSSNSSISRPAAPPEPRPQQPQPRSPAARRAA
 FT ARSRAAAAGGMRRDPAPGFSMLLFGVSLACYSPLKSVQDQ
 FT AYKAPVVVEGKVQGLVPAGGSSSNSTREPPASGRVALVKVL
 FT DKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPL
 FT VFKTAFAPLDTNGKNLKKEVGKILCTDC -> MSESRRRGR

Query Match 95.6%; Score 1505; DB 1; Length 850;
 Best Local Similarity 98.6%; Pred. No. 7.6e-113;
 Matches 285; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
 |||||
 Db 93 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 152
 Qy 61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 120
 |||||
 Db 153 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAF 212
 Qy 121 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
 |||||
 Db 213 PLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQGTGQVGEKQSLKCEAAAGNPQPSYRWFK 272
 Qy 181 DGKELNRSRDIRIKYGNRKNLSRLQFNKVVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
 |||||

Db 273 DGKELNRSRDIRIKYGNRKNRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 332

Qy 241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRG 289
 ||| : |||

Db 333 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 381

RESULT 3

NRG2_RAT

ID NRG2_RAT STANDARD; PRT; 868 AA.

AC O35569; O35073; O35570; O35571; O35572;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)

DE (Neural-and thymus-derived activator for ERBB kinases) (NTAK)].

GN NRG2 OR NTAK.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., SEQUENCE OF 128-162, AND ALTERNATIVE SPLICING.

RX MEDLINE=98006324; PubMed=9348101;

RA Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,

RA Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,

RA Ishiguro H.;

RT "A novel brain-derived member of the epidermal growth factor family

RT that interacts with ErbB3 and ErbB4.";

RL J. Biochem. 122:675-680(1997).

RN [2]

RP SEQUENCE OF 109-868 FROM N.A. (ISOFORMS 6 AND 7).

RC TISSUE=Cerebellum;

RX MEDLINE=97311397; PubMed=9168114;

RA Chang H., Riese D.J. II, Gilbert W., Stern D.F., McMahon U.J.;

RT "Ligands for ErbB-family receptors encoded by a neuregulin-like

RT gene.";

RL Nature 387:509-512(1997).

CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase

CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,

CC resulting in ligand-stimulated tyrosine phosphorylation and

CC activation of the ERBB receptors. May also promote the

CC heterodimerization with the EGF receptor.

CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS

CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE

CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=7;

CC Comment=Additional isoforms seem to exist. The alpha-type and

CC beta-type differ in the EGF-LIKE domain;

CC Name=1; Synonyms=NTAK-alpha1;

CC IsoId=O35569-1; Sequence=Displayed;

CC Name=2; Synonyms=NTAK-alpha2A;

CC IsoId=O35569-2; Sequence=VSP_003471;

CC Name=3; Synonyms=NTAK-alpha2B, NTAK-alpha2-1P;

CC IsoId=O35569-3; Sequence=VSP_003466, VSP_003471;

CC Name=4; Synonyms=NTAK-beta;

```

CC      IsoId=O35569-4; Sequence=VSP_003470;
CC      Name=5; Synonyms=NTAK-gamma;
CC      IsoId=O35569-5; Sequence=VSP_003467, VSP_003468;
CC      Name=6; Synonyms=NRG2-alpha;
CC      IsoId=O35569-6; Sequence=VSP_003472, VSP_003473;
CC      Name=7; Synonyms=NRG2-beta;
CC      IsoId=O35569-7; Sequence=VSP_003465, VSP_003469;
CC      -!- TISSUE SPECIFICITY: Expressed in most parts of the brain,
CC          especially the olfactory bulb and cerebellum where it localizes in
CC          granule and Purkinje cells. In the hippocampus, found in the
CC          granule cells of the dentate gyrus. In the basal forebrain, found
CC          in the cholinergic cells. In the hindbrain, weakly detectable in
CC          the motor trigeminal nucleus. Not detected in the hypothalamus.
CC          Also found in the liver and in the thymus. Not detected in heart,
CC          adrenal gland, or testis.
CC      -!- DEVELOPMENTAL STAGE: In the embryo, expressed in the brain of
CC          11.5 embryos where it is found in the telencephalon, but not in
CC          the hindbrain. Not found in the heart. In the adult, found in
CC          brain and thymus.
CC      -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
CC          of trafficking and proteolytic processing. Regulation of the
CC          proteolytic processing involves initial intracellular domain
CC          dimerization (By similarity).
CC      -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
CC          domain (By similarity).
CC      -!- PTM: Proteolytic cleavage close to the plasma membrane on the
CC          external face leads to the release of the soluble growth factor
CC          form (By similarity).
CC      -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC          similarity).
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D89995; BAA23344.1; -.
DR      EMBL; D89996; BAA23345.1; -.
DR      EMBL; D89997; BAA23346.1; -.
DR      EMBL; D89998; BAA23347.1; -.
DR      EMBL; AB001576; BAA23348.1; -.
DR      PIR; JC5701; JC5701.
DR      PIR; JC5702; JC5702.
DR      HSSP; Q12784; 1HRE.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      InterPro; IPR002154; Neuregulin.
DR      Pfam; PF00008; EGF; 1.
DR      Pfam; PF00047; ig; 1.

```

DR	Pfam; PF02158; Neuregulin; 1.
DR	SMART; SM00181; EGF; 1.
DR	SMART; SM00408; IGc2; 1.
DR	PROSITE; PS00022; EGF_1; 1.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS50026; EGF_3; 1.
DR	PROSITE; PS50835; IG_LIKE; 1.
KW	Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW	Transmembrane; Multigene family; Alternative splicing.
FT	PROPEP 1 127
FT	CHAIN 128 868 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.
FT	CHAIN 128 428 NEUREGULIN-2.
FT	DOMAIN 128 429 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 430 450 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT	DOMAIN 451 868 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 253 348 IG-LIKE C2-TYPE.
FT	DOMAIN 346 356 SER/THR-RICH.
FT	DOMAIN 357 398 EGF-LIKE.
FT	DOMAIN 22 32 POLY-SER.
FT	DOMAIN 35 45 POLY-SER.
FT	DOMAIN 56 59 POLY-THR.
FT	DOMAIN 103 106 POLY-ALA.
FT	DOMAIN 739 745 POLY-PRO.
FT	DISULFID 273 327 BY SIMILARITY.
FT	DISULFID 361 375 BY SIMILARITY.
FT	DISULFID 369 386 BY SIMILARITY.
FT	DISULFID 388 397 BY SIMILARITY.
FT	CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC 1 108 Missing (in isoform 7).
FT	/FTId=VSP_003465.
FT	VARSPPLIC 220 222 PLV -> FFF (in isoform 3).
FT	/FTId=VSP_003466.
FT	VARSPPLIC 388 388 C -> G (in isoform 5).
FT	/FTId=VSP_003467.
FT	VARSPPLIC 389 868 Missing (in isoform 5).
FT	/FTId=VSP_003468.
FT	VARSPPLIC 390 412 NGFFGQRCLEKLPLRLYMPDPKQ -> VGYTGDRCCQFAMV
FT	NFS (in isoform 7).
FT	/FTId=VSP_003469.
FT	VARSPPLIC 390 421 NGFFGQRCLEKLPLRLYMPDPKQKHLGFELKE -> VGYTG
FT	DRCQQFAMVNFSK (in isoform 4).
FT	/FTId=VSP_003470.
FT	VARSPPLIC 414 421 Missing (in isoform 2 and isoform 3).
FT	/FTId=VSP_003471.
FT	VARSPPLIC 414 439 HLGFELEKEAEELYQKRVLTITGICVA -> SVLWDTPGTGV
FT	SSSQWSTSPSTLDLN (in isoform 6).
FT	/FTId=VSP_003472.
FT	VARSPPLIC 440 868 Missing (in isoform 6).
FT	/FTId=VSP_003473.
FT	CONFLICT 117 117 S -> F (IN REF. 2).
FT	CONFLICT 724 724 R -> H (IN REF. 2).
SQ	SEQUENCE 868 AA; 93776 MW; 3C7D4D94DBE64DE2 CRC64;

Query Match 93.4%; Score 1470; DB 1; Length 868;
 Best Local Similarity 95.8%; Pred. No. 4.9e-110;
 Matches 277; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

```

Qy      1 MRRDPAPGFSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREP 60
        ||||||| |||||||||||||||||||||||||||||||||||||||
Db      109 MRRDPAPGSSMLLFGVSLACYSPLKSVQDQAYKAPVVVEGKVQGLAPAGGSSSNSTREP 168

Qy      61 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQPLVFKTAFE 120
        ||||||| |||||||||||||||||||||||||||||||||||||||
Db      169 PASGRVALVKVLDKWPLRSGGLQREQVISVGSCAPLERNQRYIFFLEPTEQPLVFKTAFE 228

Qy      121 PLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFK 180
        |:| |||:|||||||||||||||||||||:|||||||||
Db      229 PVDPNGKNIKKEVGKILCTDCATRPKLKKMKSQTGEVGEKQSLKCEAAAGNPQPSYRWFK 288

Qy      181 DGKELNRSRDIRIKYGNRKNRSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVS 240
        ||||||| |||||||||||||||||||||||||||||||||||:|||||
Db      289 DGKELNRSRDIRIKYGNRKNRSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGLHVNSVS 348

Qy      241 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPVGYTGDRG 289
        ||||||| ||||||||||||||||||||||| |: |||
Db      349 TTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRC 397

```

RESULT 4

NRG1_CHICK

```

ID  NRG1_CHICK  STANDARD;  PRT;  602 AA.
AC  Q05199; O73750; O73751; O73752;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1
DE  (Acetylcholine receptor inducing activity) (ARIA)].
GN  NRG1 OR ARIA.
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.
RC  STRAIN=White leghorn; TISSUE=Brain;
RX  MEDLINE=93201602; PubMed=8453670;
RA  Falls D.L., Rosen K.M., Corfas G., Lane W.S., Fischbach G.D.;
RT  "ARIA, a protein that stimulates acetylcholine receptor synthesis, is
RT  a member of the neu ligand family.";
RL  Cell 72:801-815(1993).
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).
RC  TISSUE=Brain, and Spinal cord;
RX  MEDLINE=98150951; PubMed=9491987;
RA  Yang X., Kuo Y., Devay P., Yu C., Role L.;
RT  "A cysteine-rich isoform of neuregulin controls the level of
RT  expression of neuronal nicotinic receptor channels during
RT  synaptogenesis.";
RL  Neuron 20:255-270(1998).

```

CC -!- FUNCTION: Direct ligand for the ERBB tyrosine kinase receptors.
CC The multiple isoforms perform diverse functions: Cystein-rich
CC domain containing isoforms (isoforms 2-4) probably regulate the
CC expression of nicotinic acetylcholine receptors at developing
CC interneuronal synapses. The Ig-NRG isoform is required for the
CC initial induction and/or maintenance of the mature levels of
CC acetylcholine receptors at neuromuscular synapses.
CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a
CC proteolytically released soluble growth factor form. The membrane-
CC bound form does not seem to be active (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=ARIA, IG-NRG;
CC IsoId=Q05199-1; Sequence=Displayed;
CC Note=Contains an Ig-like domain;
CC Name=2; Synonyms=CRD-NRG-BETA1A;
CC IsoId=Q05199-2; Sequence=VSP_003445;
CC Note=The EGF-like domain is replaced by a Cysteine-rich domain
CC (CRD);
CC Name=3; Synonyms=CRD-NRG-BETA2A;
CC IsoId=Q05199-3; Sequence=VSP_003445, VSP_003446;
CC Note=The EGF-like domain is replaced by a Cysteine-rich domain
CC (CRD);
CC Name=4; Synonyms=CRD-NRG-BETA2B;
CC IsoId=Q05199-4; Sequence=VSP_003445, VSP_003446, VSP_003447,
CC VSP_003448;
CC Note=The EGF-like domain is replaced by a Cysteine-rich domain
CC (CRD);
CC -!- DEVELOPMENTAL STAGE: Isoforms 2-4 are detected at embryonic day 4
CC (ED4) in both visceral and somatic motor neurons of spinal cord
CC and is highest at ED6. Isoform 1 is not expressed until ED 6 in
CC spinal cord. At ED 11 both isoforms display comparable levels.
CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
CC of trafficking and proteolytic processing. Regulation of the
CC proteolytic processing involves initial intracellular domain
CC dimerization (By similarity).
CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
CC domain.
CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the
CC external face leads to the release of the soluble growth factor
CC form.
CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC similarity).
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC -!- SIMILARITY: Belongs to the neuregulin family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L11264; AAA49037.1; -.

DR EMBL; AF045654; AAC05670.1; -.
 DR EMBL; AF045655; AAC05671.1; -.
 DR EMBL; AF045656; AAC05672.1; -.
 DR PIR; A45769; A45769.
 DR HSSP; Q12784; 1HRE.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002154; Neuregulin.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF02158; Neuregulin; 1.
 DR PRINTS; PR01089; NEUREGULIN.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 KW Transmembrane; Alternative splicing.
 FT CHAIN 1 602 PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.
 FT CHAIN 1 205 NEUREGULIN-1.
 FT DOMAIN 1 206 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 207 229 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
 FT DOMAIN 230 602 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 29 123 IG-LIKE C2-TYPE.
 FT DOMAIN 125 136 SER/THR-RICH.
 FT DOMAIN 137 181 EGF-LIKE.
 FT DISULFID 49 105 BY SIMILARITY.
 FT DISULFID 141 155 BY SIMILARITY.
 FT DISULFID 149 169 BY SIMILARITY.
 FT DISULFID 171 180 BY SIMILARITY.
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 127 MWATSEGPLQYSLAPTQTDVNSSYNTVPPKLKEMKNQEVAV
 FT GQKLVLRCETTSEYPALRFKWLKNGKEITKKNRPENVKIPK
 FT KQKKYSELHIYRATLADAGEYACRVSSKLGNSTKASVIIT
 FT DTNA -> MSEVGTETFPSPSAQLSPDASLGGLPAENMPG
 FT PHREDSRVPGVAGLASTCCVCLEAERLKGCLNSEKICIPI
 FT LACLLSLCLCIAGLKWVFVDKIFEYDSPTHLDPGRIGQDPR
 FT STVDPTALSAWVPSEVYASPFPIPSLESKA EVTQT DSSLV
 FT PSRPFLQPSLYNRILDVGLWSSATPSLSPSSLEPTTASQAQ
 FT ATETNLQTAPKLS (in isoform 2, isoform 3
 FT and isoform 4).
 FT /FTId=VSP_003445.
 FT VARSPLIC 191 198 Missing (in isoform 3 and isoform 4).
 FT /FTId=VSP_003446.
 FT VARSPLIC 388 405 VSAMTTPARMSPVDFHTP -> HTPPTSLLLAGKVSLRVS
 FT (in isoform 4).
 FT /FTId=VSP_003447.
 FT VARSPLIC 406 602 Missing (in isoform 4).
 FT /FTId=VSP_003448.
 SQ SEQUENCE 602 AA; 67453 MW; 4183C0E56CE5D346 CRC64;

Query Match 19.4%; Score 305.5; DB 1; Length 602;

Best Local Similarity 33.0%; Pred. No. 5.7e-17;
Matches 65; Conservative 37; Mismatches 74; Indels 21; Gaps 5;

```
Qy      109 TEQPLVFKTAFAPLDTNGKNLKEVGKILCTDCATRPKLKKMKSQTGQVGEKQSLKCEAA 168
          :| || : |   | |           :   ||||:||:|   ||:| |:||
Db      5  SEGPLQYSLAPTQTDVNS-----SYNTVPPKLKEMKNQEVAVGQKLVLRCETT 52

Qy      169 AGNPQPSYRWFKDGKEL---NRSRDIRIKYGNRKN SRLQFNKVKVEDAGEYVCEAENIL 225
          : |   ::| |:| |:   ||   ::|   :| | |   : : |||| |   : |
Db      53 SEYPALRFKWLKNGKEITKKNRPENVKIP-KKQKKYSELHIYRATLADAGEYACRVSSKL 111

Qy      226 GKDTVGRGLYVNSVSTTLSSWSG--HARKCNETAKSYCVNGGVCYYIEGI---NQLSCKC 280
          | |: : : : : | :| :| | ||:   |:| |||| | | : : :   |:|
Db      112 GNDSTKASVIITDTNATSTTTGTSHLTCDIKQKAFQVNGGECYMKDLNPPRYLCRC 171

Qy      281 PVGYTGDRCCQFAMVNF 297
          | :||| | | : | :|
Db      172 PNEFTGDRCCQNYVMASF 188
```

RESULT 5

NRG1_XENLA

ID NRG1_XENLA STANDARD; PRT; 677 AA.
AC O93383; Q9W6N0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1].
GN NRG1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA1), AND ALTERNATIVE SPLICING.
RX MEDLINE=98352126; PubMed=9685585;
RA Yang J.F., Zhou H., Pun S., Ip N.Y., Peng H.B., Tsim K.W.K.;
RT "Cloning of cDNAs encoding xenopus neuregulin: expression in myotomal
RT muscle during embryo development.";
RL Brain Res. Mol. Brain Res. 58:59-73(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM CRD).
RX MEDLINE=99316087; PubMed=10383827;
RA Yang J.F., Zhou H., Choi R.C., Ip N.Y., Peng H.B., Tsim K.W.K.;
RT "A cysteine-rich form of Xenopus neuregulin induces the expression of
RT acetylcholine receptors in cultured myotubes.";
RL Mol. Cell. Neurosci. 13:415-429(1999).
CC -!- FUNCTION: Direct ligand for the ERBB tyrosine kinase receptors.
CC Induces expression of acetylcholine receptor in synaptic nuclei.
CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a
CC proteolytically released soluble growth factor form. The membrane-
CC bound form does not seem to be active (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist. Isoforms have alpha
CC or beta-type EGF-like domains;

```

CC      Name=Alpha1;
CC      IsoId=O93383-1; Sequence=Displayed;
CC      Name=CRD; Synonyms=CRD-NRG1, Cysteine rich domain;
CC      IsoId=O93383-2; Sequence=VSP_003449, VSP_003450;
CC      -!- TISSUE SPECIFICITY: Isoform alpha1 is expressed in brain and
CC          muscle. Isoform CRD is expressed in brain and spinal cord, but at
CC          very low level in muscle.
CC      -!- DEVELOPMENTAL STAGE: Strong expression in developing brain and
CC          spinal cord of the embryo. Also expressed in the myotomal muscle.
CC      -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
CC          of trafficking and proteolytic processing. Regulation of the
CC          proteolytic processing involves initial intracellular domain
CC          dimerization (By similarity).
CC      -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
CC          domain.
CC      -!- PTM: Proteolytic cleavage close to the plasma membrane on the
CC          external face leads to the release of the soluble growth factor
CC          form.
CC      -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC          similarity).
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF076618; AAC26804.1; -.
DR      EMBL; AF142632; AAD33893.1; -.
DR      HSSP; Q12784; 1HRE.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      InterPro; IPR002154; Neuregulin.
DR      Pfam; PF00008; EGF; 1.
DR      Pfam; PF00047; ig; 1.
DR      Pfam; PF02158; Neuregulin; 1.
DR      PRINTS; PR01089; NEUREGULIN.
DR      SMART; SM00181; EGF; 1.
DR      SMART; SM00408; IGc2; 1.
DR      PROSITE; PS00022; EGF_1; 1.
DR      PROSITE; PS01186; EGF_2; 1.
DR      PROSITE; PS50026; EGF_3; 1.
DR      PROSITE; PS50835; IG_LIKE; 1.
KW      Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
KW      Transmembrane; Alternative splicing.
FT      CHAIN      1      259      NEUREGULIN ALPHA1 (BY SIMILARITY).
FT      CHAIN      1      677      PRO-NEUREGULIN ALPHA1, MEMBRANE-BOUND
FT                                FORM (BY SIMILARITY).
FT      DOMAIN      1      260      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM    261      280      INTERNAL SIGNAL SEQUENCE (POTENTIAL).

```

FT	DOMAIN	281	677	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	37	132	IG-LIKE C2-TYPE.
FT	DOMAIN	188	232	EGF-LIKE.
FT	DISULFID	57	116	BY SIMILARITY.
FT	DISULFID	192	206	BY SIMILARITY.
FT	DISULFID	200	220	BY SIMILARITY.
FT	DISULFID	222	231	BY SIMILARITY.
FT	DOMAIN	1	25	LYS-RICH.
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	130	130	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1	136	MAEKKKVKEGKGRKGKGGKDRKGKKAEGSDQGAAASPGLKE
FT				IKTQSVQEGKKLVLCQAVSEQPSLKFRWFKGEKEIGAKNK
FT				PDSKPEHIKIRGKKKSSELQISKASSADNGEYKCMVSNQLG
FT				NDTVTVNVTIVPK -> MSEDTAEGLQNCSEQSSDPPSAE
FT				LQNEESMPETQDEEETHGITGLAITCCVCLEADRLRICLN
FT				SEKICIIPIACLIISLCLCIAGLKWVFDKIFEYDSPHTLD
FT				PGHRGQDLILYTDTPSTLVPSSVRTLPVVIPTTDSKAAVT
FT				FKFGTSLLPTE (in isoform CRD).
FT				/FTid=VSP_003449.
FT	VARSPLIC	223	252	KPGFTGARCTETDPLRVVRSEKHLGIEFME -> PNEFTGD
FT				RCQNYVMASFYK (in isoform CRD).
FT				/FTid=VSP_003450.
SQ	SEQUENCE	677 AA;	75794 MW;	49279E8F5BAE396F CRC64;

Query Match 19.0%; Score 299.5; DB 1; Length 677;
 Best Local Similarity 33.8%; Pred. No. 2e-16;
 Matches 74; Conservative 24; Mismatches 68; Indels 53; Gaps 5;

Qy	126	GKNLKKEVGKIL---CTDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDG	182
		:: :: :: : :	
Db	15	GKGKKDRKGKKAEGSDQGAAASPGLKEIKTQSVQEGKKLVLCQAVSEQPSLKFRWFKGE	74
Qy	183	KELNR-----SRDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTV-----	230
		: :: : ::	
Db	75	KEIGAKNKPDSKPEHIKIRGKKKSSELQISKASSADNGEYKCMVSNQLGNDTVTVNVTIV	134
Qy	231	-----RGRLYVNSVSTTL-----SSWSGHARKCNE	255
		: :: : ::	
Db	135	PKPTYNHLLLMKIYLVKTVSVEKSVEPSTLNLLESQKEVIFATTKRGDTTAGPGHLIKCSD	194
Qy	256	TAKSYCVNGGVCYIIEGI---NQLSCKCPVGYTGDRCCQ	291
		: : :: :	
Db	195	KEKTYCVNGGECYVLNGITSSNQFMCKCKPGFTGARCTE	233

RESULT 6

NRG1_RAT

ID	NRG1_RAT	STANDARD;	PRT;	662 AA.
AC	P43322; P43323; P43324; P43325; P43326; P43327; P43328;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1 (Neu			
DE	differentiation factor) (Heregulin) (HRG) (Acetylcholine receptor			
DE	inducing activity) (ARIA) (Sensory and motor neuron-derived factor)			
DE	(Glial growth factor)].			
GN	NRG1 OR NDF.			

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94158863; PubMed=7509448;
 RA Wen D., Suggs S.V., Karunakaran D., Liu N., Cupples R.L., Luo Y.,
 RA Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,
 RA Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,
 RA Koski R.A., Yarden Y.;
 RT "Structural and functional aspects of the multiplicity of Neu
 RT differentiation factors.";
 RL Mol. Cell. Biol. 14:1909-1919(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA2C/NDF44), AND PARTIAL SEQUENCE.
 RC TISSUE=Fibroblast;
 RX MEDLINE=92257596; PubMed=1349853;
 RA Wen D., Peles E., Cupples R., Suggs S.V., Bacus S.S., Luo Y.,
 RA Trail G., Hu S., Silbiger S.M., Levy R.B., Koski R.A., Lu H.S.,
 RA Yarden Y.;
 RT "Neu differentiation factor: a transmembrane glycoprotein containing
 RT an EGF domain and an immunoglobulin homology unit.";
 RL Cell 69:559-572(1992).
 RN [3]
 RP SEQUENCE OF 14-36.
 RX MEDLINE=92208945; PubMed=1348215;
 RA Peles E., Bacus S.S., Koski R.A., Lu H.S., Wen D., Ogden S.G.,
 RA Levy R.B., Yarden Y.;
 RT "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein
 RT that induces differentiation of mammary tumor cells.";
 RL Cell 69:205-216(1992).
 RN [4]
 RP REGULATION OF PROCESSING (ISOFORM ALPHA2C/NDF44).
 RX MEDLINE=99069430; PubMed=9852099;
 RA Liu X., Hwang H., Cao L., Wen D., Liu N., Graham R.M., Zhou M.;
 RT "Release of the neuregulin functional polypeptide requires its
 RT cytoplasmic tail.";
 RL J. Biol. Chem. 273:34335-34340(1998).
 RN [5]
 RP INTERACTION WITH LIMK1.
 RX MEDLINE=98352096; PubMed=9685409;
 RA Wang J.Y., Frenzel K.E., Wen D., Falls D.L.;
 RT "Transmembrane neuregulins interact with LIM kinase 1, a cytoplasmic
 RT protein kinase implicated in development of visuospatial cognition.";
 RL J. Biol. Chem. 273:20525-20534(1998).
 CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase
 CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,
 CC resulting in ligand-stimulated tyrosine phosphorylation and
 CC activation of the ERBB receptors. The multiple isoforms perform
 CC diverse functions such as inducing growth and differentiation of
 CC epithelial, glial, neuronal, and skeletal muscle cells; inducing
 CC expression of acetylcholine receptor in synaptic vesicles during
 CC the formation of the neuromuscular junction; stimulating
 CC lobuloalveolar budding and milk production in the mammary gland
 CC and inducing differentiation of mammary tumor cells; stimulating

CC Schwann cell proliferation; implication in the development of the
 CC myocardium such as trabeculation of the developing heart (By
 CC similarity).
 CC -!- SUBUNIT: The cytoplasmic domain interacts with the LIM domain
 CC region of LIMK1.
 CC -!- SUBCELLULAR LOCATION: Exists as a type I membrane protein and as a
 CC proteolytically released soluble growth factor form. The membrane-
 CC bound form does not seem to be active.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=8;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Beta4; Synonyms=NDF42A;
 CC IsoId=P43322-1; Sequence=Displayed;
 CC Name=Alpha2A; Synonyms=NDF38;
 CC IsoId=P43322-2; Sequence=VSP_003436;
 CC Name=Alpha2B; Synonyms=NDF19;
 CC IsoId=P43322-3; Sequence=VSP_003436, VSP_003443, VSP_003444;
 CC Name=Alpha2C; Synonyms=NDF44;
 CC IsoId=P43322-4; Sequence=VSP_003436, VSP_003442;
 CC Name=Beta1;
 CC IsoId=P43322-5; Sequence=VSP_003437;
 CC Name=Beta2; Synonyms=NDF40;
 CC IsoId=P43322-6; Sequence=VSP_003440, VSP_003441;
 CC Name=Beta2A; Synonyms=NDF22;
 CC IsoId=P43322-7; Sequence=VSP_003440;
 CC Name=Beta3; Synonyms=NDF4;
 CC IsoId=P43322-8; Sequence=VSP_003438, VSP_003439;
 CC -!- TISSUE SPECIFICITY: Widely expressed. Most tissues contain alpha2A
 CC and alpha2B isoforms. Alpha2 and beta2 are the predominant forms
 CC in mesenchymal and nonneuronal organs. Beta1 is enriched in brain
 CC and spinal cord, but not in muscle and heart. Alpha2C is highly
 CC expressed in spinal cord, moderately in lung, brain, ovary, and
 CC stomach, in low amounts in the kidney, skin and heart and not
 CC detected in the liver, spleen, and placenta.
 CC -!- DOMAIN: The cytoplasmic domain may be involved in the regulation of
 CC trafficking and proteolytic processing. Regulation of the
 CC proteolytic processing involves initial intracellular domain
 CC dimerization.
 CC -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
 CC domain.
 CC -!- PTM: Proteolytic cleavage close to the plasma membrane on the
 CC external face leads to the release of the soluble growth factor
 CC form.
 CC -!- PTM: Extensive glycosylation precedes the proteolytic cleavage.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
 CC -!- SIMILARITY: Belongs to the neuregulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U02315; AAA19940.1; -.

DR EMBL; U02316; AAA19941.1; -.
 DR EMBL; U02317; AAA19942.1; -.
 DR EMBL; U02318; AAA19943.1; -.
 DR EMBL; U02319; AAA19944.1; -.
 DR EMBL; U02320; AAA19945.1; -.
 DR EMBL; U02321; AAA19946.1; -.
 DR EMBL; U02322; AAA19947.1; -.
 DR EMBL; U02323; AAA19948.1; -.
 DR EMBL; U02324; AAA19949.1; -.
 DR EMBL; M92430; -; NOT_ANNOTATED_CDS.
 DR PIR; I61718; I61718.
 DR PIR; I61719; I61719.
 DR PIR; I61722; I61722.
 DR HSSP; Q12784; LHRE.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR002154; Neuregulin.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF02158; Neuregulin; 1.
 DR PRINTS; PR01089; NEUREGULIN.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;
 KW Transmembrane; Multigene family; Alternative splicing.
 FT PROPEP 1 13
 FT CHAIN 14 662 PRO-NEUREGULIN-1, MEMBRANE-BOUND FORM.
 FT CHAIN 14 264 NEUREGULIN-1.
 FT DOMAIN 14 265 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 266 288 INTERNAL SIGNAL SEQUENCE (POTENTIAL).
 FT DOMAIN 289 662 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 37 128 IG-LIKE C2-TYPE.
 FT DOMAIN 165 177 SER/THR-RICH.
 FT DOMAIN 178 222 EGF-LIKE.
 FT DISULFID 57 112
 FT DISULFID 182 196 BY SIMILARITY.
 FT DISULFID 190 210 BY SIMILARITY.
 FT DISULFID 212 221 BY SIMILARITY.
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 213 256 PNEFTGDRCQNYVMASFYMTSRRKRQETEKPLERKLDHSLV
 FT KES -> QPGFTGARCTENVPMKVQTQE (in isoform
 FT Alpha2A, isoform Alpha2B and isoform
 FT Alpha2C).
 FT /FTId=VSP_003436.
 FT VARSPLIC 231 257 MTSRRKRQETEKPLERKLDHSLVKESK -> KHLGIEFME
 FT (in isoform Beta1).
 FT /FTId=VSP_003437.
 FT VARSPLIC 231 241 MTSRRKRQETE -> STSTPFLSLPE (in isoform
 FT Beta3).

FT				/FTId=VSP_003438.
FT	VARSPPLIC	242	662	Missing (in isoform Beta3).
FT				/FTId=VSP_003439.
FT	VARSPPLIC	231	256	Missing (in isoform Beta2 and isoform Beta2A).
FT				/FTId=VSP_003440.
FT	VARSPPLIC	325	330	PPENVQ -> RVRTRG (in isoform Beta2).
FT				/FTId=VSP_003441.
FT	VARSPPLIC	446	662	Missing (in isoform Alpha2C).
FT				/FTId=VSP_003442.
FT	VARSPPLIC	446	484	YVSAMTTPARMSPVDFHTPSSPKSPPEMSPPVSSMTVS
FT				-> HNLIAELRRNKAYRSKCMQIQLSATHLRPSSITHLGFI
FT				L (in isoform Alpha2B).
FT				/FTId=VSP_003443.
FT	VARSPPLIC	485	662	Missing (in isoform Alpha2B).
FT				/FTId=VSP_003444.
FT	CONFLICT	90	90	K -> N (IN REF. 2).
FT	CONFLICT	137	137	T -> I (IN REF. 2; AA SEQUENCE).
FT	CONFLICT	208	208	Y -> S (IN REF. 2).

Query Match 19.0%; Score 299; DB 1; Length 662;
 Best Local Similarity 33.8%; Pred. No. 2.1e-16;
 Matches 67; Conservative 34; Mismatches 53; Indels 44; Gaps 6;

Qy	142	ATRPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGLNRS---RDIRIKYGNG	198
		: : : :: : : : :	
Db	34	ALPPRLKEMKSQESAAGSKVLRCETSSEYSSLRFKWFKNGNELNRKNKPNIKIQKPG	93
Qy	199	RKNSRLQFNKVKVEDAGEYVCEAENILGKDTVGRGL-----YV----	236
		: : : : : : : : :	
Db	94	K--SELRINKASLADSGEYMCKVISKLGNDASANITIVESNEFITGMPASTETAYVSSE	151
Qy	237	-----NSVSTTLSSWSG--HARKCNETAKSYCVNGGVVCYYIEGINQLS---CK	279
		: : : : : : : : : :	
Db	152	SPIRISVSTEGANTSSSTSTSTTGTSHLIKCAEKEKTFVCVNGGECFTVKDLNPSRYLCK	211
Qy	280	CPVGYTGDRCCQFAMVNF	297
		: : :	
Db	212	CPNEFTGDRCCQNYVMASF	229

RESULT 7

NRG1_HUMAN

ID NRG1_HUMAN STANDARD; PRT; 639 AA.
 AC Q02297; O14667; P98202; Q02298; Q02299; Q07110; Q07111; Q12779;
 AC Q12780; Q12781; Q12782; Q12783; Q12784; Q9UPE3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Pro-neuregulin-1 precursor (Pro-NRG1) [Contains: Neuregulin-1 (Neu
 DE differentiation factor) (Heregulin) (HRG) (Breast cancer cell
 DE differentiation factor p45) (Acetylcholine receptor inducing activity)
 DE (ARIA) (Sensory and motor neuron-derived factor) (Glial growth
 DE factor)].
 GN NRG1 OR HGL OR NDF OR HRGA OR GGF OR SMDF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 6; 7 AND 8), AND PARTIAL
 RP SEQUENCE.
 RX MEDLINE=92271253; PubMed=1350381;
 RA Holmes W.E., Sliwkowski M.X., Akita R.W., Henzel W.J., Lee J.,
 RA Park J.W., Yansura D., Abadi N., Raab H., Lewis G.D., Shepard H.M.,
 RA Kuang W.-J., Wood W.I., Goeddel D.V., Vandlen R.L.;
 RT "Identification of heregulin, a specific activator of p185erbB2.";
 RL Science 256:1205-1210(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 6; 7 AND 8).
 RC TISSUE=Kidney adenocarcinoma, and Pituitary;
 RX MEDLINE=94158863; PubMed=7509448;
 RA Wen D., Suggs S.V., Karunakaran D., Liu N., Cupples R.L., Luo Y.,
 RA Janssen A.M., Ben-Baruch N., Trollinger D.B., Jacobsen V.L.,
 RA Meng S.-Y., Lu H.S., Hu S., Chang D., Yang W., Yanigahara D.,
 RA Koski R.A., Yarden Y.;
 RT "Structural and functional aspects of the multiplicity of Neu
 RT differentiation factors.";
 RL Mol. Cell. Biol. 14:1909-1919(1994).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=92208945; PubMed=1348215;
 RA Peles E., Bacus S.S., Koski R.A., Lu H.S., Wen D., Ogden S.G.,
 RA Levy R.B., Yarden Y.;
 RT "Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein
 RT that induces differentiation of mammary tumor cells.";
 RL Cell 69:205-216(1992).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS 8 AND 9).
 RC TISSUE=Brain;
 RX MEDLINE=93205115; PubMed=8096067;
 RA Marchionni M.A., Goodearl A.D.J., Chen M.S., Bermingham-McDonogh O.,
 RA Kirk C., Hendricks M., Danehy F., Misumi D., Sudhalter J.,
 RA Kobayashi K., Wroblewski D., Lynch C., Baldassarre M., Hiles I.,
 RA Davis J.B., Hsuan J.J., Totty N.F., Otsu M., McBurney R.N.,
 RA Waterfield M.D., Stroobant P., Gwynne D.;
 RT "Glial growth factors are alternatively spliced erbB2 ligands
 RT expressed in the nervous system.";
 RL Nature 362:312-318(1993).
 RN [5]
 RP SEQUENCE FROM N.A. OF GAMMA-HEREGULIN FUSION PROTEIN.
 RC TISSUE=Breast cancer;
 RX MEDLINE=97472144; PubMed=9333014;
 RA Schaefer G., Fitzpatrick V.D., Sliwkowski M.X.;
 RT "Gamma-heregulin: a novel heregulin isoform that is an autocrine
 RT growth factor for the human breast cancer cell line, MDA-MB-175.";
 RL Oncogene 15:1385-1394(1997).
 RN [6]
 RP SEQUENCE OF 1-210 FROM N.A.
 RA Schoumacher F., Herzer S., Flury N., Kueng W., Mueller H.,
 RA Eppenberger U.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
 RN [7]
 RP SEQUENCE OF 19-27.

RX MEDLINE=93366731; PubMed=7689552;
 RA Culouscou J.-M., Plowman G.D., Carlton G.W., Green J.M., Shoyab M.;
 RT "Characterization of a breast cancer cell differentiation factor that
 RT specifically activates the HER4/p180erbB4 receptor.";
 RL J. Biol. Chem. 268:18407-18410(1993).
 RN [8]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=99455251; PubMed=10523851;
 RA Wang X.-Z., Jolicoeur E.M., Conte N., Chaffanet M., Zhang Y.,
 RA Mozziconacci M.-J., Feiner H., Birnbaum D., Pebusque M.-J., Ron D.;
 RT "Gamma-heregulin is the product of a chromosomal translocation fusing
 RT the DOC4 and HGL/NRG1 genes in the MDA-MB-175 breast cancer cell
 RT line.";
 RL Oncogene 18:5718-5721(1999).
 RN [9]
 RP CHROMOSOMAL TRANSLOCATION.
 RX MEDLINE=20065180; PubMed=10597312;
 RA Liu X., Baker E., Eyre H.J., Sutherland G.R., Zhou M.;
 RT "Gamma-heregulin: a fusion gene of DOC-4 and neuregulin-1 derived from
 RT a chromosome translocation.";
 RL Oncogene 18:7110-7114(1999).
 RN [10]
 RP STRUCTURE BY NMR OF 175-241 (ISOFORM 1).
 RX MEDLINE=94341264; PubMed=8062828;
 RA Nagata K., Kohda D., Hatanaka H., Ichikawa S., Matsuda S.,
 RA Yamamoto T., Suzuki A., Inagaki F.;
 RT "Solution structure of the epidermal growth factor-like domain of
 RT heregulin-alpha, a ligand for p180erbB-4.";
 RL EMBO J. 13:3517-3523(1994).
 CC -!- FUNCTION: Direct ligand for ERBB3 and ERBB4 tyrosine kinase
 CC receptors. Concomitantly recruits ERBB1 and ERBB2 coreceptors,
 CC resulting in ligand-stimulated tyrosine phosphorylation and
 CC activation of the ERBB receptors. The multiple isoforms perform
 CC diverse functions such as inducing growth and differentiation of
 CC epithelial, glial, neuronal, and skeletal muscle cells; inducing
 CC expression of acetylcholine receptor in synaptic vesicles during
 CC the formation of the neuromuscular junction; stimulating
 CC lobuloalveolar budding and milk production in the mammary gland
 CC and inducing differentiation of mammary tumor cells; stimulating
 CC Schwann cell proliferation; implication in the development of the
 CC myocardium such as trabeculation of the developing heart.
 CC -!- SUBUNIT: The cytoplasmic domain interacts with the LIM domain
 CC region of LIMK1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Exists as an type I membrane protein and as
 CC a proteolytically released soluble growth factor form. The
 CC membrane-bound form does not seem to be active. The secreted
 CC isoform 9 has a signal peptide. The isoform 8 may be nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=9;
 CC Comment=Additional isoforms seem to exist. Isoforms have been
 CC classified as type I NRGS (isoforms with an Ig domain and a
 CC glycosylation domain, isoforms 1-8), type II NRGS (isoforms with
 CC an Ig domain but no glycosylation domain, isoform 9) and type
 CC III NRGS (isoforms with a Cys-rich domain, isoform 10). All
 CC these isoforms perform distinct tissue-specific functions;
 CC Name=1; Synonyms=Alpha;
 CC IsoId=Q02297-1; Sequence=Displayed;

```

CC      Name=2; Synonyms=AlphaA;
CC      IsoId=Q02297-2; Sequence=VSP_003431;
CC      Name=3; Synonyms=Alpha2B;
CC      IsoId=Q02297-3; Sequence=VSP_003434, VSP_003435;
CC      Name=4; Synonyms=Alpha3;
CC      IsoId=Q02297-4; Sequence=VSP_003432, VSP_003433;
CC      Name=6; Synonyms=Beta1, Beta1A;
CC      IsoId=Q02297-6; Sequence=VSP_003428;
CC      Name=7; Synonyms=Beta2;
CC      IsoId=Q02297-7; Sequence=VSP_003427;
CC      Name=8; Synonyms=Beta3, GGFHFB1;
CC      IsoId=Q02297-8; Sequence=VSP_003429, VSP_003430;
CC      Name=9; Synonyms=GGF2, GGFHPP2;
CC      IsoId=Q02297-9; Sequence=VSP_003425, VSP_003426, VSP_003429,
CC                               VSP_003430;
CC      Name=10; Synonyms=SMDF;
CC      IsoId=Q15491-1; Sequence=External;
CC      -!- TISSUE SPECIFICITY: Type I isoforms are the predominant forms
CC          expressed in the endocardium. Isoform alpha is expressed in
CC          breast, ovary, testis, prostate, heart, skeletal muscle, lung,
CC          placenta liver, kidney, salivary gland, small intestine and brain,
CC          but not in uterus, stomach, pancreas, and spleen. Isoform 3 is the
CC          predominant form in mesenchymal cells and in nonneuronal organs,
CC          whereas isoform 5 is the major neuronal form. Isoform 8 is
CC          expressed in spinal cord and brain. Isoform 9 is the major form in
CC          skeletal muscle cells; in the nervous system it is expressed in
CC          spinal cord and brain. Also detected in adult heart, placenta,
CC          lung, liver, kidney, and pancreas.
CC      -!- DEVELOPMENTAL STAGE: Detectable at early embryonic ages.
CC      -!- DOMAIN: The cytoplasmic domain may be involved in the regulation
CC          of trafficking and proteolytic processing. Regulation of the
CC          proteolytic processing involves initial intracellular domain
CC          dimerization (By similarity).
CC      -!- DOMAIN: ERBB receptor binding is elicited entirely by the EGF-like
CC          domain.
CC      -!- PTM: Proteolytic cleavage close to the plasma membrane on the
CC          external face leads to the release of the soluble growth factor
CC          form.
CC      -!- PTM: Extensive glycosylation precedes the proteolytic cleavage (By
CC          similarity).
CC      -!- DISEASE: Involved in a rare t(8;11) chromosomal translocation that
CC          fuses the 5'end of ODZ4 to NRG1 (isoform 8). The product of this
CC          translocation was first thought to be an alternatively spliced
CC          isoform, called gamma-heregulin. Gamma-heregulin is a soluble
CC          activating ligand for the ERBB2-ERBB3 receptor complex and acts as
CC          an autocrine growth factor in a specific breast cancer cell line
CC          (MDA-MB-175). Not detected in breast carcinoma samples, including
CC          ductal, lobular, medullary, and mucinous histological types,
CC          neither in other breast cancer cell lines.
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M94165; AAA58638.1; -.
DR EMBL; M94166; AAA58639.1; -.
DR EMBL; M94167; AAA58640.1; -.
DR EMBL; M94168; AAA58641.1; -.
DR EMBL; U02325; AAA19950.1; -.
DR EMBL; U02326; AAA19951.1; -.
DR EMBL; U02327; AAA19952.1; -.
DR EMBL; U02328; AAA19953.1; -.
DR EMBL; U02329; AAA19954.1; -.
DR EMBL; U02330; AAA19955.1; -.
DR EMBL; L12260; AAB59622.1; -.

Query Match 17.8%; Score 280.5; DB 1; Length 639;
Best Local Similarity 32.1%; Pred. No. 6.1e-15;
Matches 69; Conservative 35; Mismatches 60; Indels 51; Gaps 7;

Qy 126 GKNLKKEVGKILCTDCAT-----RPKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRW 178
|| ||| | : | | :||:|||| | | |:|| :: :||
Db 10 GKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCESTSEYSSSLRFRKW 69

Qy 179 FKDGKELNRS---RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRGL- 234
||:| |||| ::||:| | : | | : || : ||||:| : || | : :
Db 70 FKNGNELNRKNKPQNIKIQKKPGK--SELRINKASLADSGEYMCKVISKLGNDSSASANIT 127

Qy 235 -----YV-----NSVSTTLSSWSG--HARKCNETAKS 259
|| | : |:| :| :| | || | | :
Db 128 IVESNEIITGMPASTEGAYVSSESPIRISVSTEGANTSSSTSTTTGTSHLVKCAEKEKT 187

Qy 260 YCVNGGVCCYYIEGINQLS---CKCPVGYTGDRCCQ 291
:||||| |: :: :: | ||| |:|| || :
Db 188 FCVNGGECFMVKDLNPSRYLCKCQPGFTGARCTE 222

RESULT 8

VEIN_DROME

ID VEIN_DROME STANDARD; PRT; 623 AA.
AC Q94918; Q9VRQ3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Vein protein precursor (Epidermal growth factor-like protein)
DE (Defective dorsal discs protein).
GN VN OR DDD OR CG10491.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Embryo, and Imaginal disks;
RX MEDLINE=96421972; PubMed=8824589;
RA Schnepf B.C., Grumbling G.B., Donaldson T.D., Simcox A.A.;

RT "Vein is a novel component in the Drosophila epidermal growth factor
 RT receptor pathway with similarity to the neuregulins.";
 RL Genes Dev. 10:2302-2313(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";

RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 CC -!- FUNCTION: Ligand for the EGF receptor. Seems to play a role in
 CC the global proliferation of wing disc cells and the larval
 CC patterning. Shows a strong synergistic genetic interaction with
 CC spi, suggesting a molecular interdependence. Required for the
 CC development of interveins cells.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q94918-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q94918-2; Sequence=VSP_001419;
 CC -!- DEVELOPMENTAL STAGE: Expressed in blastoderm embryos in two
 CC ventrolateral stripes that are brought to the midline as
 CC gastrulation proceeds. In the germ-band retraction stage,
 CC expression is seen in the CNS and epidermis. At late blastoderm,
 CC expression is localized in the anlagen of the amnioserosa.
 CC Expression in the head, clypeolabrum, maxillary and labial lobes,
 CC and around the stomodeum throughout embryo development. In late
 CC embryos, expression decays in all ectodermal cells and appears in
 CC the segmental muscles and the gut wall. In the larva, expression
 CC occurs in the dorsal metathoracic disc, the eye-antennal disc and
 CC the ventral thoracic disc. Found in the intervein in the pupa.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67935; AAC47293.1; -.
 DR EMBL; AE003564; AAF50739.2; -.
 DR FlyBase; FBgn0003984; vn.
 DR GO; GO:0005576; C:extracellular; NAS.
 DR GO; GO:0005154; F:epidermal growth factor receptor binding; IMP.
 DR GO; GO:0007173; P:EGF receptor signaling pathway; IMP.
 DR GO; GO:0007477; P:notum morphogenesis; IMP.
 DR GO; GO:0045742; P:positive regulation of EGF receptor signali. . .; NAS.
 DR GO; GO:0007476; P:wing morphogenesis; IMP.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; ig; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; FALSE_NEG.
 DR PROSITE; PS50026; EGF_3; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW EGF-like domain; Glycoprotein; Immunoglobulin domain; Growth factor;

KW Developmental protein; Alternative splicing; Signal.

FT	SIGNAL	1	40	POTENTIAL.
FT	CHAIN	41	623	VEIN PROTEIN.
FT	DOMAIN	457	542	IG-LIKE C2-TYPE.
FT	DOMAIN	561	599	EGF-LIKE.
FT	DOMAIN	43	53	POLY-GLN.
FT	DOMAIN	87	96	POLY-SER.
FT	DOMAIN	138	314	GLN-RICH.
FT	DISULFID	478	531	BY SIMILARITY.
FT	CARBOHYD	76	76	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	252	252	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	350	350	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	381	381	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	424	424	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	449	449	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	521	521	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	574	574	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSP LIC	609	623	FVAIYQGIHTLNNDY -> SSPESCKNYQGGYY (in isoform 2).
FT				/FTid=VSP_001419.
FT	CONFLICT	149	149	MISSING (IN REF. 1).
FT	CONFLICT	220	220	S -> T (IN REF. 1).
FT	CONFLICT	494	494	E -> D (IN REF. 1).
SQ	SEQUENCE	623 AA;	71697 MW;	AFD2724D5C1F56C8 CRC64;

Query Match 13.2%; Score 208.5; DB 1; Length 623;
 Best Local Similarity 25.8%; Pred. No. 3.4e-09;
 Matches 79; Conservative 42; Mismatches 104; Indels 81; Gaps 15;

Qy	32	AYKAPVVVEG-----KVQGLVPAGGSSSNSTREPPA	62
		: : :: : :	
Db	329	AFAAPTQVFQGVFKSMSADRRVNFSA TMKVEKVYKQQHDLQLPTLVRLQFALSNSSGECD-	387
Qy	63	SGRVALVKVLDKWPLRSGG-LQREQVISV GSCVPLERNQRYIFFLEPTE-----QP-	112
		: :: : : : ::	
Db	388	---IYRERLMPRGMLRSGNDLQQASDIS-----YMMFVQQTNPNGNFTILGQPM	432
Qy	113	----LVFKTAFAPLDTNGKNLKEVGKILCTDCATRPKLKMKMSQTGQVGEKQSLKCEAA	168
		: : : : : :	
Db	433	RVTHLVVEAVETAVSEN-YTQNAEVT KIF-----SKPSKAI IKH-----GKKLRIVCE-V	480
Qy	169	AGNPQPSYRWFKDGKELNRSRDIRIKYGN GRKNSRL---QFNKVKVEDAGEYVCEAENIL	225
		: : : :: : :: :	
Db	481	SGQPPPKVTWFKDEKSINRKRNI-YQFKH HKRRSELIVRSFN--SSSDAGRYE C RAKNKA	537
Qy	226	GKDTVGRGLYVNSVSTTL-SSWSGHARKCNETAKSYCVNGGV CYIIEGINQLSCKCPVGY	284
		: : : : : :	
Db	538	SKAI AKRRIMIKASPVHFPTDRSASGIPC N---FDYCFHNGTCRMIPDINEVYCRCPTEY	594
Qy	285	TGDR CQ	290
		: :	
Db	595	FGNRCE	600

RESULT 9
 JAM2_HUMAN

ID JAM2_HUMAN STANDARD; PRT; 298 AA.
 AC P57087;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Junctional adhesion molecule 2 precursor (Vascular endothelial
 DE junction-associated molecule) (VE-JAM).
 GN JAM2 OR VEJAM OR C21ORF43.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Vascular endothelial cells;
 RX MEDLINE=20317114; PubMed=10779521;
 RA Palmeri D., van Zante A., Huang C.C., Hemmerich S., Rosen S.D.;
 RT "Vascular endothelial junction-associated molecule, a novel member of
 RT the immunoglobulin superfamily, is localized to intercellular
 RT boundaries of endothelial cells.";
 RL J. Biol. Chem. 275:19139-19145(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=20507930; PubMed=10945976;
 RA Cunningham S.A., Arrate M.P., Rodriguez J.M., Bjercke R.J.,
 RA Vanderslice P., Morris A.P., Brock T.A.;
 RT "A novel protein with homology to the junctional adhesion molecule:
 RT Characterization of leukocyte interactions.";
 RL J. Biol. Chem. 275:34750-34756(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE PROCESSES OF LYMPHOCYTE HOMING TO
 CC SECONDARY LYMPHOID ORGANS.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).

CC -!- TISSUE SPECIFICITY: PROMINENTLY EXPRESSED ON HIGH ENDOTHELIAL
 CC VENULES BUT IS ALSO PRESENT ON THE ENDOTHELIA OF OTHER VESSELS.
 CC LOCALIZED TO THE INTERCELLULAR BOUNDARIES OF HIGH ENDOTHELIAL
 CC CELLS.
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:1-3(2001);
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1652492186_g.htm".

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF255910; AAF81223.1; -.
 DR EMBL; AY016009; AAG49022.1; -.
 DR EMBL; BC017779; AAH17779.1; -.
 DR Genew; HGNC:14686; JAM2.
 DR MIM; 606870; -.
 DR GO; GO:0005887; C:integral to plasma membrane; NAS.
 DR GO; GO:0016337; P:cell-cell adhesion; NAS.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00047; ig; 2.
 DR SMART; SM00408; IGc2; 1.
 DR PROSITE; PS50835; IG_LIKE; 2.
 KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 298 JUNCTIONAL ADHESION MOLECULE 2.
 FT DOMAIN 21 238 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 239 259 POTENTIAL.
 FT DOMAIN 260 298 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 32 127 IG-LIKE V-TYPE.
 FT DOMAIN 134 238 IG-LIKE C2-TYPE.
 FT DISULFID 50 109 POTENTIAL.
 FT DISULFID 155 214 POTENTIAL.
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 298 AA; 33207 MW; CA78E518E22DCAEE CRC64;

Query Match 9.8%; Score 155; DB 1; Length 298;
 Best Local Similarity 27.7%; Pred. No. 2.8e-05;
 Matches 56; Conservative 24; Mismatches 86; Indels 36; Gaps 8;

Qy 96 LERNQRYIFFLEPTEQPLVFKTAFAPLDTN--GKNL-KKEVGKILCTDCATRPKLKKMKS 152
 | | : : : : : | | : : : | | : : :
 Db 66 LGRSVSFVYYQQTLQGD--FKNRAEMIDFNIRIKNVTRSDAGKYRCEVSAPSEQGQNLEE 123
 Qy 153 QTGQV-----GEKQSLKCEAAAGNPQPSYRWFKDGKELNR-----S 188
 | : | : : | | | | | | | |
 Db 124 DVTLEVLVAPAVPSCEVPSSALSGTVVELRCQDKEGNPAPEYTWFKDGIRLLENPRLGS 183


```

Qy      189 RDIRIKYGNRKNRSLQFNKVKVEDAGEYVCEAENILGKDTVGRG-RLVNSVSTTLSSWS 247
      :      |      |      |||| |      | ||| ||| | :|      | |: |: ::      |
Db      184 QSTNSSYTMNTKTGTTLQFNTVSKLDTGEYSCEARNSVGYRRCPGKRMQVDDLNI-----S 238

Qy      248 GHARKCNETAKSYCVNG-GVCY 268
      |      |      | || |||
Db      239 GIIAAVVVVALVISVCGLGVCY 260

```

RESULT 10

SMDF_HUMAN

```

ID      SMDF_HUMAN      STANDARD;      PRT;      296 AA.
AC      Q15491;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Neuregulin-1, sensory and motor neuron-derived factor isoform.
GN      NRG1 OR HGL OR NDF OR HRGA OR GGF OR SMDF.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain stem, and Cerebellum;
RX      MEDLINE=95301541; PubMed=7782315;
RA      Ho W.-H., Armanini M.P., Nuijens A., Phillips H.S., Osheroff P.L.;
RT      "Sensory and motor neuron-derived factor. A novel heregulin variant
RT      highly expressed in sensory and motor neurons.";
RL      J. Biol. Chem. 270:14523-14532(1995).
CC      -!- FUNCTION: The isoform SMDF may play a role in motor and sensory
CC      neuron development.
CC      -!- SUBCELLULAR LOCATION: Secreted. May possess an internal uncleaved
CC      signal sequence.
CC      -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=10;
CC      Comment=Additional isoforms seem to exist. Isoforms have been
CC      classified as type I NRGs (isoforms with an IG domain and a
CC      glycosylation domain, alpha and beta), Type II NRGs (isoforms
CC      with an IG domain but no glycosylation domain - GGF2) and type
CC      III NRGs (isoforms with a cys-rich domain - SMDF). All these
CC      isoforms perform distinct tissue-specific functions;
CC      Name=10; Synonyms=SMDF;
CC      IsoId=Q15491-1; Sequence=Displayed;
CC      Name=1; Synonyms=Alpha;
CC      IsoId=Q02297-1; Sequence=External;
CC      Name=2; Synonyms=Alpha1A;
CC      IsoId=Q02297-2; Sequence=External;
CC      Name=3; Synonyms=Alpha2B;
CC      IsoId=Q02297-3; Sequence=External;
CC      Name=4; Synonyms=Alpha3;
CC      IsoId=Q02297-4; Sequence=External;
CC      Name=5; Synonyms=Beta1;
CC      IsoId=Q02297-5; Sequence=External;
CC      Name=6; Synonyms=Beta1A;
CC      IsoId=Q02297-6; Sequence=External;
CC      Name=7; Synonyms=Beta2;

```

```

CC      IsoId=Q02297-7; Sequence=External;
CC      Name=8; Synonyms=Beta3, GGFHFB1;
CC      IsoId=Q02297-8; Sequence=External;
CC      Name=9; Synonyms=GGF2, GGFHPP2;
CC      IsoId=Q02297-9; Sequence=External;
CC      -!- TISSUE SPECIFICITY: Expressed in nervous system: spinal cord motor
CC          neurons, dorsal root ganglion neurons, and brain. Predominant
CC          isoform expressed in sensory and motor neurons. Not detected in
CC          adult heart, placenta, lung, liver, skeletal muscle, kidney, and
CC          pancreas. Not expressed in fetal lung, liver and kidney.
CC      -!- DEVELOPMENTAL STAGE: Highly expressed in developing spinal motor
CC          neurons and in developing cranial nerve nuclei. Expression is
CC          maintained only in both adult motor neurons and dorsal root
CC          ganglion neurons.
CC      -!- SIMILARITY: Contains 1 EGF-like domain.
CC      -!- SIMILARITY: Belongs to the neuregulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L41827; AAC41764.1; -.
DR      PIR; A56943; A56943.
DR      HSSP; Q12784; 1HRE.
DR      Genew; HGNC:7997; NRG1.
DR      MIM; 142445; -.
DR      GO; GO:0016020; C:membrane; NAS.
DR      GO; GO:0007399; P:neurogenesis; NAS.
DR      InterPro; IPR006209; EGF_like.
DR      InterPro; IPR006210; IEGF.
DR      Pfam; PF00008; EGF; 1.
DR      SMART; SM00181; EGF; 1.
DR      PROSITE; PS00022; EGF_1; 1.
DR      PROSITE; PS01186; EGF_2; FALSE_NEG.
DR      PROSITE; PS50026; EGF_3; 1.
KW      Growth factor; EGF-like domain; Transmembrane; Multigene family;
KW      Alternative splicing.
FT      TRANSMEM      76      100      INTERNAL SIGNAL SEQUENCE (POTENTIAL).
FT      DOMAIN        58      91      CYS-RICH.
FT      DOMAIN        211     232     SER/THR-RICH.
FT      DOMAIN        233     277     EGF-LIKE.
FT      DISULFID      237     251     BY SIMILARITY.
FT      DISULFID      245     265     BY SIMILARITY.
FT      DISULFID      267     276     BY SIMILARITY.
SQ      SEQUENCE      296 AA;  31685 MW;  8D41743217F7EB02 CRC64;

```

```

Query Match          9.7%;  Score 152;  DB 1;  Length 296;
Best Local Similarity 22.0%;  Pred. No. 4.8e-05;
Matches 71; Conservative 49; Mismatches 116; Indels 86; Gaps 16;

```

```

Qy      21 YSPSLKSVQDQAYKAP---VVVEGKVQGLVPAGGSSSNSTREPPASGRVALV----KVLD 73
      ||| : | : :| : : : || | : | | | :
Db      4 YSPDMSEVAAERSSSPSTQLSADPSLDGLPAAEDMPEPQTEDGRTPLGLVGLAVPCCACLE 63

```

Qy 74 KWPLRSGGLQREQV-----ISVGSCVPLERNQRYIF-----FLEPT-----E 110
 || || | : : : : : : : : : : : : : : : :
 Db 64 AERLR-GCLNSEKICIVPILACLVSCLCI---AGLKWVFDKIFEYDSPHTLDPGGLGQ 119

Qy 111 QPLVFKTAFAPLDTNGKNLKKEVGKILCTDCATRPKLKKMKSQTGQV---GEKQSLKCEA 167
 | :
 Db 120 DPII-----SLDATAAS-----AVWVSSEAYTSPVSRQSESEVQVTVQGDKAVVSFEP 168

Qy 168 AAGNPQPSYRWF-----KDGKELNRSRDIRIKYGNRKNRSLQFNKVKVEDAGEYV 218
 : | | | | | :
 Db 169 SAA-PTPKNRIFAFLPSTAPSFPSPTRNPEVR----TPKSATQPQTETNLQTAPK-- 221

Qy 219 CEAENILGKDTVGRGLYVNSVSTTLSSWSGHARKCNETAKSYCVNGGVCYYIEGINQLS- 277
 | | | | : | | | | :
 Db 222 -----LSTSTSTTGTS---HLVKCAEKEKTFCVNGGECFMVKDLSNPSR 262

Qy 278 --CKCPVGYTGDRCCQFAMVNF 297
 | | | | : | | | | : | : |
 Db 263 YLCKCPNEFTGDRCCQNYVMASF 284

RESULT 11

EGF_MOUSE

ID EGF_MOUSE STANDARD; PRT; 1217 AA.
 AC P01132;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor].
 GN EGF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83223630; PubMed=6602382;
 RA Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Fong N.M.,
 RA Selby M., Rutter W.J., Bell G.I.;
 RT "Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins."
 RL Science 221:236-240(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83219309; PubMed=6304537;
 RA Gray A., Dull T.J., Ullrich A.;
 RT "Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor."
 RL Nature 303:722-725(1983).
 RN [3]
 RP SEQUENCE OF 977-1029.
 RX MEDLINE=73048516; PubMed=4636327;
 RA Savage C.R. Jr., Inagami T., Cohen S.;
 RT "The primary structure of epidermal growth factor."
 RL J. Biol. Chem. 247:7612-7621(1972).

RN [4]
 RP DISULFIDE BONDS.
 RX MEDLINE=74025498; PubMed=4750422;
 RA Savage C.R. Jr., Hash J.H., Cohen S.;
 RT "Epidermal growth factor. Location of disulfide bonds.";
 RL J. Biol. Chem. 248:7669-7672(1973).
 RN [5]
 RP STRUCTURE BY NMR OF 977-1029.
 RX MEDLINE=92118798; PubMed=1731873;
 RA Montelione G.T., Wuethrich K., Burgess A.W., Nice E.C., Wagner G.,
 RA Gibson K.D., Scheraga H.A.;
 RT "Solution structure of murine epidermal growth factor determined by
 RT NMR spectroscopy and refined by energy minimization with
 RT restraints.";
 RL Biochemistry 31:236-249(1992).
 RN [6]
 RP STRUCTURE BY NMR OF 977-1029.
 RX MEDLINE=93075811; PubMed=1445923;
 RA Kohda D., Inagaki F.;
 RT "Three-dimensional nuclear magnetic resonance structures of mouse
 RT epidermal growth factor in acidic and physiological pH solutions.";
 RL Biochemistry 31:11928-11939(1992).
 RN [7]
 RP STRUCTURE BY NMR OF 980-1024.
 RX MEDLINE=99180407; PubMed=10082370;
 RA Barnham K.J., Torres A.M., Alewood D., Alewood P.F., Domagala T.,
 RA Nice E.C., Norton R.S.;
 RT "Role of the 6-20 disulfide bridge in the structure and activity of
 RT epidermal growth factor.";
 RL Protein Sci. 7:1738-1749(1998).
 CC -!- FUNCTION: The growth factor stimulates the growth of various
 CC epidermal and epithelial tissues in vivo and in vitro and of some
 CC fibroblasts in cell culture.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: Contains 9 EGF-like domains.
 CC -!- CAUTION: Ref.2 sequence differs from that shown in positions 1134
 CC to 1168 due to a frameshift.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J00380; AAA37539.1; -.
 DR EMBL; V00741; CAA24115.1; ALT_FRAME.
 DR EMBL; V00741; CAA24116.1; -.
 DR PIR; A94272; EGMSMG.
 DR PDB; 1EGF; 31-JAN-94.
 DR PDB; 3EGF; 31-JAN-94.
 DR PDB; 1EPG; 31-JAN-94.
 DR PDB; 1EPH; 31-JAN-94.
 DR PDB; 1EPI; 31-JAN-94.
 DR PDB; 1EPJ; 31-JAN-94.
 DR PDB; 1A3P; 29-JUL-98.

DR PDB; 1GK5; 08-AUG-02.
 DR MGD; MGI:95290; Egf.
 DR InterPro; IPR000152; Asx_hydroxyl_S.
 DR InterPro; IPR001336; EGF_1.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR000033; Ldl_receptor_rep.
 DR Pfam; PF00008; EGF; 8.
 DR Pfam; PF00058; ldl_recept_b; 7.
 DR PRINTS; PR00009; EGFTGF.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00135; LY; 7.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 6.
 DR PROSITE; PS50026; EGF_3; 5.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
 KW Signal; 3D-structure.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 1217 PRO-EPIDERMAL GROWTH FACTOR.
 FT CHAIN 977 1029 EPIDERMAL GROWTH FACTOR.
 FT DOMAIN 29 1038 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1039 1058 POTENTIAL.
 FT DOMAIN 1059 1217 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 327 361 EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN 362 402 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 403 443 EGF-LIKE 3.
 FT DOMAIN 441 483 EGF-LIKE 4.
 FT DOMAIN 747 787 EGF-LIKE 5.
 FT DOMAIN 838 876 EGF-LIKE 6.
 FT DOMAIN 877 918 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 919 959 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 978 1019 EGF-LIKE 9.
 FT DISULFID 366 377 BY SIMILARITY.
 FT DISULFID 373 386 BY SIMILARITY.
 FT DISULFID 388 401 BY SIMILARITY.
 FT DISULFID 407 418 BY SIMILARITY.
 FT DISULFID 414 427 BY SIMILARITY.
 FT DISULFID 429 442 BY SIMILARITY.
 FT DISULFID 445 457 BY SIMILARITY.
 FT DISULFID 453 467 BY SIMILARITY.
 FT DISULFID 469 482 BY SIMILARITY.
 FT DISULFID 751 762 BY SIMILARITY.
 FT DISULFID 758 771 BY SIMILARITY.
 FT DISULFID 773 786 BY SIMILARITY.
 FT DISULFID 842 853 BY SIMILARITY.
 FT DISULFID 847 862 BY SIMILARITY.
 FT DISULFID 864 875 BY SIMILARITY.
 FT DISULFID 881 895 BY SIMILARITY.
 FT DISULFID 888 904 BY SIMILARITY.
 FT DISULFID 906 917 BY SIMILARITY.
 FT DISULFID 923 936 BY SIMILARITY.
 FT DISULFID 930 945 BY SIMILARITY.
 FT DISULFID 947 958 BY SIMILARITY.
 FT DISULFID 982 996
 FT DISULFID 990 1007

FT	DISULFID	1009	1018	
FT	DOMAIN	1024	1029	NOT REQUIRED FOR FULL BIOLOGICAL
FT				ACTIVITY.
FT	CARBOHYD	111	111	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	810	810	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	944	944	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	790	790	D -> Y (IN REF. 2).
FT	CONFLICT	1048	1048	A -> S (IN REF. 2).
FT	TURN	983	984	
FT	STRAND	995	998	
FT	TURN	1000	1002	
FT	STRAND	1005	1008	
FT	TURN	1011	1012	
FT	STRAND	1013	1014	
FT	TURN	1015	1018	
FT	STRAND	1020	1021	
SQ	SEQUENCE	1217 AA; 133144 MW; A9C7F3D512F82873 CRC64;		

Query Match 9.5%; Score 150; DB 1; Length 1217;
 Best Local Similarity 23.4%; Pred. No. 0.00037;
 Matches 64; Conservative 31; Mismatches 104; Indels 74; Gaps 11;

Qy	65	RVALVKVLD-----KWPLRSG---GLQREQVISVGSCVPLERNQRYIFFLEPTEQPLV	114
		: : : : :	
Db	774	REGFVKAWDGKMCLPQDYPILSGENADLSKE-VTSLSNSTQAE-----VPDDDGTE	823
Qy	115	FKTAFAPLDTNGKNLKEVGKILC---TDCATRPKLKKMKSQTGQVGEKQSLKCEAAAGN	171
		: : : : : : : :	
Db	824	SSTLVAEIMVSGMNYEDDCGPGGCGSHARCVS-----DGETAECQCLKGFARDGN	873
Qy	172	-----PQPSYRWFKDGKELNRSRDIRIKYGNRKNRSRLQFNKVKVEDAGE	216
		: : : : :	
Db	874	LCSDIDECVLARSDCPSTSSRC-----INTEGGYVCRCSGEYEGDGISCFDIDECQGA	927
Qy	217	YVCEAEN-----ILGKDTVGRGLYVNSVSTTLSSWSGHARK-----CNETA	257
		: : : : : : : :	
Db	928	HNC-AENAACTNTEGGYNCTCAGRPSSPGRSCPDSTAPSLLEDGHHLDRNSYPGCPSSY	986
Qy	258	KSYCVNNGVCYYIEGINQLSCKCPVGYTGDRQC	290
		: : : : : :	
Db	987	DGYCLNGGVCMHIESLDSYTCNCVIGYSGDRQC	1019

RESULT 12

DSCA_HUMAN

ID DSCA_HUMAN STANDARD; PRT; 2012 AA.
 AC O60469; O60468;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Down syndrome cell adhesion molecule precursor (CHD2).
 GN DSCAM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=98087574; PubMed=9426258;
 RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
 RA Lyons G.E., Korenberg J.R.;
 RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
 RT Down syndrome region and is involved in the development of the
 RT nervous system.";
 RL Hum. Mol. Genet. 7:227-237(1998).
 RN [2]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RX MEDLINE=20384934; PubMed=10925149;
 RA Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
 RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
 RT intercellular adhesion.";
 RL Brain Res. Mol. Brain Res. 79:118-126(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 CC -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
 CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
 CC NERVOUS SYSTEM DEVELOPMENT.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
 CC SHORT ISOFORM MAY BE SECRETED.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long; Synonyms=CHD2-52;
 CC IsoId=O60469-1; Sequence=Displayed;
 CC Name=Short; Synonyms=CHD2-42;
 CC IsoId=O60469-2; Sequence=VSP_002502, VSP_002503;
 CC -!- TISSUE SPECIFICITY: Primarily expressed in brain.
 CC -!- SIMILARITY: Contains 10 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 6 fibronectin type III domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF023450; AAC17967.1; -.
 DR EMBL; AF023449; AAC17966.1; -.
 DR EMBL; AF217525; AAF27525.1; -.
 DR EMBL; AL163283; CAB90464.1; -.
 DR EMBL; AL163282; CAB90436.1; -.
 DR EMBL; AL163281; CAB90444.1; -.
 DR Genew; HGNC:3039; DSCAM.
 DR MIM; 602523; -.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0007399; P:neurogenesis; TAS.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00041; fn3; 6.
 DR Pfam; PF00047; ig; 9.
 DR SMART; SM00060; FN3; 6.
 DR SMART; SM00408; IGc2; 7.
 DR PROSITE; PS50835; IG_LIKE; 9.
 KW Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
 KW Transmembrane; Alternative splicing.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 2012 DOWN SYNDROME CELL ADHESION MOLECULE.
 FT DOMAIN 18 1595 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1596 1616 POTENTIAL.
 FT DOMAIN 1617 2012 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 39 129 IG-LIKE C2-TYPE 1.
 FT DOMAIN 125 216 IG-LIKE C2-TYPE 2.
 FT DOMAIN 225 305 IG-LIKE C2-TYPE 3.
 FT DOMAIN 313 401 IG-LIKE C2-TYPE 4.
 FT DOMAIN 407 500 IG-LIKE C2-TYPE 5.
 FT DOMAIN 504 592 IG-LIKE C2-TYPE 6.
 FT DOMAIN 596 685 IG-LIKE C2-TYPE 7.
 FT DOMAIN 690 783 IG-LIKE C2-TYPE 8.
 FT DOMAIN 787 883 IG-LIKE C2-TYPE 9.
 FT DOMAIN 885 972 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 984 1076 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 1088 1177 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 1189 1273 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 1285 1377 IG-LIKE C2-TYPE 10.
 FT DOMAIN 1380 1463 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 1477 1562 FIBRONECTIN TYPE-III 6.
 FT DISULFID 46 102 BY SIMILARITY.
 FT DISULFID 145 197 BY SIMILARITY.
 FT DISULFID 246 293 BY SIMILARITY.
 FT DISULFID 335 385 BY SIMILARITY.
 FT DISULFID 428 484 BY SIMILARITY.
 FT DISULFID 525 575 BY SIMILARITY.
 FT DISULFID 617 669 BY SIMILARITY.
 FT DISULFID 711 766 BY SIMILARITY.
 FT DISULFID 809 865 BY SIMILARITY.
 FT DISULFID 1307 1359 BY SIMILARITY.
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	487	487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	512	512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	658	658	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	795	795	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	924	924	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1160	1160	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1271	1271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1488	1488	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSP LIC	1562	1571	NFATLNYDGS -> KEAARCKEFS (in isoform Short).
FT				/FTId=VSP_002502.
FT	VARSP LIC	1572	2012	Missing (in isoform Short).
FT				/FTId=VSP_002503.
FT	CONFLICT	1893	2012	HRPGDLIHLPPYL RMDFLN RGGPGTSRDLSLGQACLEPQK
FT				SRTLKRPTVLEPI PMEAASSASSTREGQSWQPGAVATLPQR
FT				EGAELGQA AKMSSSQESLLDSRGHLKGNNPYAKSYTLV ->
FT				IGQVTSYICLHTLEWTF C (IN REF. 1).
SQ	SEQUENCE	2012 AA;	222259 MW;	0E33CFB781A08334 CRC64;

Query Match 9.2%; Score 145.5; DB 1; Length 2012;
 Best Local Similarity 24.1%; Pred. No. 0.0015;
 Matches 77; Conservative 34; Mismatches 111; Indels 97; Gaps 15;

Qy	5	PAPGF SMLLF GVS LAC--YSPSLK-SVQDQAYKAPVVVEGKVQGLVPAGGSSSNSTREPP	61
		: : : :	
Db	87	PPSSFSTLIHDNTYYCTAENPSGKIRSQDVHIKAVL-----	REP- 125
Qy	62	ASGRVALVKVL D KWPLRSGGLQREQVISVGSCVPLERNQRYIFFL--EPTEQPLVFKTAF	119
		: : : : : : : :	
Db	126	-----YTVRVEDQKTMRGN-----VAVFKCIIPSSVEAYITVVSWEKDTVSLVSGSRF	173
Qy	120	APLDTNG---KNLKKEVGKILCTDCATRPK-----	LKKMK 151
		: : : :	
Db	174	LITSTGALYIKDVQNE DG-LYNYRCITRHRYTGETRQNSARLFVSDPANSAPSILDGFD	232
Qy	152	SQTGQVG EKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRKN-SRLQFNKVK	210
		: : : : : : : : :	
Db	233	HRKAMAGQRVELPCK-ALGHPEPDYRWLKD NMPLLS-----GRFQKTVTGLLIENIR	284
Qy	211	VEDAGEYVCEAENILGKDTVGRGRLYVNS-VSTTLSSWSGHARKCNETAKSYCVNGGVCYY	269
		: : : :	
Db	285	PSDSGSYVCEVSNRYGTAKVIGRLYVKQPLKATIS-----PRKVKSSVGS-----	329
Qy	270	IEGINQLSCKCPVGYTGDR	288
		: :	
Db	330	-----QVSLSCSVTGTEDQ	343

RESULT 13
 LAMP_RAT

ID LAMP_RAT STANDARD; PRT; 338 AA.
 AC Q62813;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Limbic system-associated membrane protein precursor (LSAMP).
 GN LSAMP OR LAMP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-49.
 RC TISSUE=Hippocampus;
 RX MEDLINE=95374785; PubMed=7646886;
 RA Pimenta A.F., Zhukareva V., Barbe M.F., Reinoso B.S., Grimley C.,
 RA Henzel W., Fischer I., Levitt P.;
 RT "The limbic system-associated membrane protein is an Ig superfamily
 RT member that mediates selective neuronal growth and axon targeting.";
 RL Neuron 15:287-297(1995).
 CC -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
 CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
 CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
 CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- TISSUE SPECIFICITY: Expressed mostly by neurons comprising limbic-
 CC associated cortical and subcortical regions that function in
 CC cognition, emotion, memory, and learning.
 CC -!- DEVELOPMENTAL STAGE: FIRST DETECTED AT E15-16, AT STAGE E20 IT IS
 CC DETECTED IN PRESUMPTIVE CORTEX, MEDIAL LIMBIC AREAS OF THE
 CC THALAMUS AND HYPOTHALAMUS. IN THE ADULT, IT IS FOUND IN
 CC HYPOTHALAMUS, PERIRHINAL CORTEX, AMYGDALA AND MEDIAL THALAMIC
 CC REGION.
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
 CC family.
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U31554; AAA86120.1; -.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00047; ig; 3.
 DR SMART; SM00408; IGc2; 2.
 DR PROSITE; PS50835; IG_LIKE; 3.
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
 KW Repeat; Signal; Lipoprotein.
 FT SIGNAL 1 28
 FT CHAIN 29 315 LIMBIC SYSTEM-ASSOCIATED MEMBRANE
 FT PROTEIN.
 FT PROPEP 316 338 REMOVED IN MATURE FORM (POTENTIAL).

FT DOMAIN 29 122 IG-LIKE C2-TYPE 1.
 FT DOMAIN 132 214 IG-LIKE C2-TYPE 2.
 FT DOMAIN 219 304 IG-LIKE C2-TYPE 3.
 FT DISULFID 53 111 POTENTIAL.
 FT DISULFID 153 197 POTENTIAL.
 FT DISULFID 239 290 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT LIPID 315 315 GPI-anchor amidated asparagine
 FT (Potential).
 SQ SEQUENCE 338 AA; 37324 MW; 0B76AFDD68A39BB6 CRC64;

Query Match 9.2%; Score 145; DB 1; Length 338;
 Best Local Similarity 23.9%; Pred. No. 0.0002;
 Matches 55; Conservative 31; Mismatches 92; Indels 52; Gaps 9;

Qy 53 SSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISVGSCVPLERNQRYIFFLEPTEQP 112
 | : || | :|| | : : :|| | : :
 Db 112 SVQTQHEPKTSQVYLIVQV----PPKISNISSDVTVNEGSNVTL-----VCMANGRPEP 161

 Qy 113 LVFKTAFAPL-----DTNGKNLKKEVGKILCTDCAT-----RPKL 147
 :: || : :|| | :: | :
 Db 162 VITWRHLTPLGREFEGEEYLEILGITREQSGKYECKAANEVSSADVQVKVTVNYPPTI 221

 Qy 148 KKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKD GKELNRSRDIRIKYGNRGRKNSRLQFN 207
 : || | : ||||| : | | : || : | : : || : |
 Db 222 TESKSNEATTGRQASLKCEASA-VPAPDFEWYRDDTRINSANGLEIKSTEGQ--SSLTVT 278

 Qy 208 KVKVEDAGEYVCEAENILG-----KDTVGRGLYVN-SVSTTLSSW 246
 | | | | | | | :|| :| | : |
 Db 279 NVTEEHYGNITCVAANKLGVTNASLVLFPGSVRG---INGSISLAVPLW 325

RESULT 14

LAMP_HUMAN

ID LAMP_HUMAN STANDARD; PRT; 338 AA.
 AC Q13449;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Limbic system-associated membrane protein precursor (LSAMP).
 GN LSAMP OR LAMP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96235133; PubMed=8666243;
 RA Pimenta A.F., Fischer I., Levitt P.;
 RT "cDNA cloning and structural analysis of the human limbic-system-

```

RT   associated membrane protein (LAMP).";
RL   Gene 170:189-195(1996).
CC   -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC       CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC       MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC       OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: Expressed on limbic neurons and fiber tracts
CC       as well as in single layers of the superior colliculus, spinal
CC       chord and cerebellum.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC       family.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U41901; AAC50569.1; -.
DR   PIR; JC4776; JC4776.
DR   Genew; HGNC:6705; LSAMP.
DR   MIM; 603241; -.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Lipoprotein.
FT   SIGNAL          1      28      POTENTIAL.
FT   CHAIN           29     315     LIMBIC SYSTEM-ASSOCIATED MEMBRANE
FT                                     PROTEIN.
FT   PROPEP          316     338     REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN           29     122     IG-LIKE C2-TYPE 1.
FT   DOMAIN          132     214     IG-LIKE C2-TYPE 2.
FT   DOMAIN          219     304     IG-LIKE C2-TYPE 3.
FT   DISULFID         53     111     POTENTIAL.
FT   DISULFID        153     197     POTENTIAL.
FT   DISULFID        239     290     POTENTIAL.
FT   CARBOHYD         40      40     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD         66      66     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        136     136     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        148     148     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        279     279     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        287     287     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        300     300     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD        315     315     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID           315     315     GPI-anchor amidated asparagine
FT                                     (Potential).
SQ   SEQUENCE      338 AA;  37308 MW;  03455F286DF5D92F CRC64;

```

Query Match 9.1%; Score 144; DB 1; Length 338;

Best Local Similarity 25.0%; Pred. No. 0.00025;
Matches 56; Conservative 34; Mismatches 94; Indels 40; Gaps 10;

```
QY      53 SSNSTREPPASGRVALVKVLDKWPLRSGGLQREQVISVGS-----CVPLERNQRYIFF-- 105
      | : || | :|| | : : : || | : | : | :
Db     112 SVQTQHEPKTSQVYLIVQV----PPKISNISSDVTVNEGSNVTILVCMANGRPEPVITWRH 167

QY     106 LEPTEQPLVEFKTAFAPL-----DTNGKNLKKKEVGKILCTDCAT-----RPKLKKMKSQ 153
      | || : : : : : : || | : : | : : ||
Db     168 LTPTGREFEGEEEEYLEILGITREQSGKYECKAANEVSSADVQVKVTVNYPPTITESKSN 227

QY     154 TGQVGEKQSLKCEAAAGNPQPSYRWFKDGGKELNRSRDIRIKYGNRKN SRLQFNKVKVED 213
      | : ||||| : | | : | : | : | : | : | : | : | :
Db     228 EATTGRQASLKCEASA-VPAPDFEWYRDDTRINSANGLEIKSTEGQ--SSLTVTNVTEEH 284

QY     214 AGEYVCEAENILG-----KDTVGRGLYVN-SVSTTLSSW 246
      | | | | | | : || : | | : | : |
Db     285 YGNYTCVAANKLGV TNASLV LFRPGSVRG---INGSISLAVPLW 325
```

RESULT 15

CEPU_CHICK

ID CEPU_CHICK STANDARD; PRT; 353 AA.
AC Q90773;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE CEPU-1 protein precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=96370549; PubMed=8774445;
RA Spaltmann F., Bruemmendorf T.;
RT "CEPU-1, a novel immunoglobulin superfamily molecule, is expressed by
RT developing cerebellar Purkinje cells."
RL J. Neurosci. 16:1770-1779(1996).
CC -!- FUNCTION: It may be a cellular address molecule specific to
CC Purkinje cells. It may represent a receptor or a subunit of a
CC receptor complex.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Minor;
CC IsoId=Q90773-1; Sequence=Displayed;
CC Name=2; Synonyms=Major;
CC IsoId=Q90773-2; Sequence=VSP_002607;
CC -!- TISSUE SPECIFICITY: Found on the dendrites, somata and axons of
CC developing Purkinje cells. Undetectable on other neurons like
CC Golgi or granule cells.
CC -!- DEVELOPMENTAL STAGE: Expressed by developing cerebellar Purkinje
CC cells. Expression coincides with the growth of the dendritic tree,
CC after Purkinje cells have finished their migration from the

CC ventricular zone (from E15 until E21). Expressed in the adult.
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
 CC family.
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z72497; CAA96578.1; -.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00047; ig; 3.
 DR SMART; SM00408; IGc2; 2.
 DR PROSITE; PS50835; IG_LIKE; 3.
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
 KW Repeat; Signal; Alternative splicing; Lipoprotein.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 330 CEPU-1 PROTEIN.
 FT PROPEP 331 353 REMOVED IN MATURE FORM (POTENTIAL).
 FT DOMAIN 37 124 IG-LIKE C2-TYPE 1.
 FT DOMAIN 134 216 IG-LIKE C2-TYPE 2.
 FT DOMAIN 220 314 IG-LIKE C2-TYPE 3.
 FT DISULFID 55 113 POTENTIAL.
 FT DISULFID 155 199 POTENTIAL.
 FT DISULFID 241 293 POTENTIAL.
 FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT LIPID 330 330 GPI-anchor amidated serine (Potential).
 FT VARSPPLIC 310 320 Missing (in isoform 2).
 FT /FTid=VSP_002607.
 SQ SEQUENCE 353 AA; 38736 MW; 2550C48591EBBBA6 CRC64;

Query Match 9.0%; Score 141.5; DB 1; Length 353;
 Best Local Similarity 36.5%; Pred. No. 0.00041;
 Matches 38; Conservative 11; Mismatches 52; Indels 3; Gaps 3;

QY 145 PKLKKMKSQTGQVGEKQSLKCEAAAGNPQPSYRWFKDGKELNRSRDIRIKYGNRKN SRL 204
 | : || ||:| | |||:| | ::|:| | | : :| | |||
 Db 221 PYISDAKSTGVPVGQKGILMCEASA-VPSADFQWYKDDKRLAEGQK-GLKVENKAFFSRL 278
 QY 205 QFNKVKVEDAGEYVCEAENILGKDTVGRGLYVNSVSTTLSSWSG 248
 | | :| | | | | | | | | | | | | :| | : | |
 Db 279 TFFNVSEQDYGNITCVASNQLGNTNASMILY-EETTTALTPWKG 321

Search completed: August 17, 2004, 14:11:18
 Job time : 9.5414 secs